U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Genetic diversity, conservation and evolutionary history of the African wild ass (*Equus africanus*): a non-invasive molecular approach

Sónia Patrícia de Melo C. Pires Rosenbom

Tese de Doutoramento apresentada à
Faculdade de Ciências da Universidade do Porto

Programa Doutoral em Biodiversidade, Genética e Evolução

2016

Genetic diversity, conservation and evolutionary history of the African wild ass (*Equus africanus*): a non-invasive molecular approach

Sónia Patrícia de Melo C. Pires Rosenbom

Programa Doutoral em Biodiversidade, Genética e Evolução
Departamento de Biologia
2016

Orientador
Albano Beja-Pereira, PhD, Investigador Principal, Professor Auxiliar Convidado
Faculdade de Ciências da Universidade do Porto

COMPETE
PROGRAMA OPERACIONAL FACTORES DE COMPETITIVIDADE

QUADRO DE REFERÊNCIA ESTRATÉGICO NACIONAL
PORTUGAL 2007.2013

UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento Regional

Declaração

Na elaboração desta dissertação e nos termos do nº2 do Artigo 8º do Decreto-lei nº 388/70 da República Portuguesa, os resultados de trabalhos já publicados foram totalmente aproveitados e fazem parte integrante de alguns capítulos desta dissertação. Em todos estes trabalhos, a candidata participou na obtenção, interpretação, análise e discussão dos resultados e na elaboração das suas formas publicadas.

Este trabalho foi realizado na íntegra no Centro em Biodiversidade e Recursos Genéticos (CIBIO) da Faculdade de Ciências da Universidade do Porto, sob a orientação do Professor Doutor Albano Beja-Pereira.

Este trabalho foi financiado por fundos do FEDER através de um projeto cofinanciado pelo programa COMPETE da EU e por fundos nacionais da Fundação para a Ciência e a Tecnologia (FCT) - FCOMP-01-0124-FEDER-019757; PTDC/BIA-BDE/64111/2006 e PTDC/BIA-BIC/118107/2010.

Statement

In this thesis, and according to number 2 of Article 8 of law number 388/70 of the Portuguese Republic, the results of published works were totally used and included in some of the chapters. In all these works, the candidate participated in obtaining, interpreting, analyzing and discussing the results and in writing the published forms.

This work was conducted in the Research Centre in Biodiversity and Genetic Resources (CIBIO) of the Science Faculty of the University of Porto, under the supervision of Professor Albano Beja-Pereira.

This work was funded by FEDER funds though a project co-funded by the EU program COMPETE and the national funds from the Portuguese Foundation for Science and Technology (FCT) – FCOMP-01-0124-FEDER-019757; PTDC/BIA-BDE/64111/2006 and PTDC/BIA-BIC/118107/2010.

To my amazing kids, Karoline and Lukas…

To my love and life partner, Kim…

You are the best part of my life…

I love you beyond words!

"And, when you want something, all the universe conspires in helping you to achieve it."

Paulo Coelho - The Alchemist

Agradecimentos / Acknowledgements

Em primeiro lugar tenho a agradecer ao meu orientador, Doutor Albano Beja-Pereira, pela oportunidade de reiniciar a minha atividade académica numa área científica bastante diversa da minha formação inicial. O seu empenho, entusiasmo, capacidade de coordenação e orientação foram imprescindíveis para a elaboração desta Tese, que resulta do trabalho colaborativo de uma vasta equipa.

À Vânia Costa devo agradecer a sua enorme paciência e dedicação na orientação de todas as tarefas laboratoriais e na discussão de resultados. Tenho que agradecer também a sua disponibilidade para ler e fazer sugestões a alguns dos artigos que compõem esta Tese. Muito obrigado por tudo!

To Jay Chen, I would like to thank for all the guidance on the phylogenetic analyses, as well as the important advices about relevant literature.

I would like to thank Dr. Patricia Moehlman, for the enormous efforts in coordinating field work in Ethiopia and Eritrea, as well as reading and correcting some of the manuscripts that are a part of this Thesis. Your work and commitment to wild equid conservation, namely the African wild ass, is truly inspiring…

To all the colleagues that conducted field work in extremely harsh conditions, in Ethiopia, Eritrea, Iran, China and Mongolia…No words are enough…Thanks to: Fanuel Kebede, Redae Teclai, Hagos Yohannes, Futsum Hagos, Jay Chen, Leili Khalatbari, Gholam Hosein Yusefi, Ablimit Abdukadir and Chamba Yangzom.

To Kefena Efa for the great cooperation in the study of domestic donkeys and horses from Ethiopia.

Tenho que agradecer a todos os meus colegas de laboratório/grupo pela enorme simpatia, ajuda e companheirismo: Rui Gomes, Rui Oliveira, Raquel Xavier, Catarina Rato, Joana Rocha, Joana Vilela, Ana Matos, João Maia, Carolina Pereira, Bruno Maia, Liliana Farelo, Pedro Coelho, Susana Freitas e Duarte Gonçalves. A vossa boa disposição ajudou-me nos dias em que nada corria bem no laboratório… Obrigado!

A todo o pessoal do CTM, nomeadamente à Susana Lopes, Sara João e Diana Castro, gostaria de agradecer o apoio na utilização das instalações e na obtenção de dados.

Gostaria também de agradecer à Sara Lemos e à Sandra Rodrigues por toda a ajuda prestada ao longo dos anos de doutoramento.

Por fim gostaria de agradecer à minha família por todo o apoio e carinho demonstrado!

Resumo

O burro selvagem Africano (*Equus africanus*) é uma das oito espécies de Equídeos sobreviventes. Esta espécie carrega a herança genética de um processo evolutivo de 55 milhões de anos, correspondente à longa linha evolutiva da família Equidae, tendo sido a única das três espécies de "burros selvagens" a ser domesticada, dando origem ao burro doméstico.

Historicamente o burro selvagem Africano encontrava-se distribuído numa vasta área geográfica do Nordeste Africano, tendo ocupado um território que iria do Sul do Egito atá ao Norte da Somália e da Etiópia. Porém, a ação humana e as alterações ambientais dos últimos séculos, reduziram substancialmente a sua distribuição, que se encontra atualmente circunscrita ao deserto do Danakil, na Eritreia e Etiópia, e mesmo nesta área, o seu habitat encontra-se bastante fragmentado.

O enorme declínio no número de indivíduos, tanto na Eritreia como na Etiópia, transformou o burro selvagem Africano numa espécie vulnerável, sendo atualmente considerado o Equídeo selvagem mais ameaçado segundo a IUCN e, por tal, é altamente passível de se extinguir num futuro próximo, se nada for feito.

Dado o atual cenário, é urgente o desenvolvimento e aplicação de medidas de conservação que impeçam a depleção da diversidade existente, comprometendo desta forma a sobrevivência

desta espécie. No entanto, para conservar, é necessário (1) avaliar parâmetros demográficos e de diversidade relevantes, nas populações sobreviventes da Etiópia e Eritreia e (2) clarificar as relações filogenéticas entre o burro selvagem Africano e as restantes espécies de burros selvagens (*Equus hemionus* e *Equus kiang*), que ainda se encontram pouco estudadas.

Finalmente, sendo o burro selvagem Africano o mais provável ancestral selvagem do burro doméstico (*E. a. asinus*) é ainda possível avaliar a contribuição da espécie selvagem no "pool genético" do burro doméstico, sendo assim possível compreender melhor os efeitos do processo de domesticação na estrutura genética e populacional da espécie domesticada em relação à sua ancestral selvagem.

A reavaliação da história evolutiva dos burros selvagens foi feita usando uma metodologia de amostragem não invasiva, através da colheita de fezes de indivíduos pertencentes às populações sobreviventes de burros selvagens Africanos e Asiáticos, numa vasta área da sua distribuição geográfica, que vai do corno da África no continente Africano, ao planalto Tibetano, no Extremo Oriente.

As sequências de ADN mitocondrial obtidas foram usadas para estimar os níveis de diversidade intraespecífica e os seus padrões de distribuição geográfica, e para reconstruir as relações filogenéticas entre espécies/subespécies do grupo dos burros selvagens.

A análise das sequências do citocromo b (Cyt *b*) de burros selvagens Africanos e Asiáticos, pertencentes às populações sobreviventes, demostrou que este grupo de indivíduos partilhou um ancestral comum há aproximadamente 2.3 milhões de anos (I.C. 95% entre 1.4 e 3.2 milhões de anos), o que é coincidente com o intervalo de tempo deduzido a partir do registo fóssil para o ancestral comum deste ramo da filogenia, suportando a hipótese de burros e cavalos selvagens terem coexistido no Norte da América, numa fase inicial da sua história evolutiva.

Os padrões filogeográficos de burros selvagens Africanos e Asiáticos revelaram-se bastante distintos, no entanto, no que diz respeito ao burro selvagem Africano, não foi encontrada estruturação geográfica dos haplótipos mitocondriais, com três dos quatro haplótipos obtidos a serem partilhados entre as populações da Etiópia e da Eritreia. Os resultados obtidos revelaram relações filogenéticas entre as espécies/subespécies de burros Asiáticos que diferem da taxonomia correntemente aceite para este grupo e requerem uma reavaliação mais exaustiva.

Os primeiros dados de genética das populações do burro selvagem Africano, foram obtidos para a diversidade genética, estruturação populacional e parâmetros demográficos, pela genotipagem de um conjunto de microssatélites autossómicos. Os valores de diversidade genética observados

e estimados apontam claramente a população da Eritreia, como a mais diversa. A análise dos padrões migratórios permitiu a deteção de migração bilateral histórica e contemporânea entre as populações da Etiópia e da Eritreia. Estes resultados são concordantes com a ausência de estruturação geográfica, encontrada através da análise do ADN mitocondrial, confirmando-se a ausência de estruturação através da análise de dados nucleares.

Haplótipos mitocondriais previamente observados em animais domésticos foram detetados em indivíduos morfologicamente identificados como burros selvagens Africanos, tanto na população da Etiópia como na Eritreia. Apesar das limitações impostas pelo reduzido número de marcadores na identificação de híbridos entre as formas doméstica e selvagem, foi possível identificar um individuo na população da Eritreia que apresentava um padrão de miscigenação consistente com um híbrido de primeira geração (F1). Os nossos resultados apontam para a existência de hibridização esporádica e limitada à população da Eritreia entre as formas doméstica e selvagem.

Devido à sua proximidade genética, o burro selvagem Africano é considerado como o mais provável ancestral do burro doméstico. A identificação de duas linhagens perfeitamente definidas (denominadas Clado 1 e Clado 2) entre os haplótipos mitocondriais de burros domésticos parece ser o resultado de dois eventos separados de domesticação. A existência de duas subespécies de burros selvagens Africanos (burro da Nubia e o burro da Somália), morfologicamente distintas, foi também documentada.

A nossa análise de haplótipos mitocondriais obtidos a partir de amostras antigas, históricas e pertencentes a Museus, de burros selvagens Africanos e burros domésticos, confirmou o papel central do burro da Nubia, na origem dos burros domésticos do Clado 1. De facto, os burros selvagens da Nubia e os burros domésticos do Clado 1 são praticamente indistinguíveis, com base na análise do ADN mitocondrial, com burros selvagens da Nubia possuindo haplótipos idênticos aos de burros domésticos do Clado 1.

A topologia do Clado 1 é consistente com um cenário no qual o burro selvagem da Nubia terá sido domesticado em diversas áreas e/ou durante um período de tempo extenso, com múltiplos recrutamentos a partir das populações selvagens.

O extenso conjunto de dados obtidos a partir de indivíduos pertencentes a populações naturais do burro selvagem da Somália, demostrou que estes são claramente distintos dos burros selvagens da Nubia e dos burros domésticos de ambos os Clados 1 e 2. A baixa variabilidade encontrada num elevado número de amostras de burros selvagens da Somália, fazem com que

a probabilidade de encontrar novas linhagens seja baixa, o que faz do burro selvagem da Somália um ancestral menos provável dos burros domésticos do Clado 2.

Presentemente, os dados arqueológicos existentes apontam para duas regiões geográficas como as mais prováveis para a origem do burro doméstico: o Nordeste Africano e a Península Arábica. Neste trabalho, avaliamos e comparamos os níveis e padrões de diversidade genética, usando dados de microssatélites, de oito populações de burros domésticos de três hipotéticos centros de origem do Nordeste Africano, Médio Oriente e Península Arábica. Apesar do Médio Oriente de ter sido sugerido em tempos como um putativo centro de origem do burro domestico, os dados arqueológicos obtidos nas últimas décadas, têm vindo a afastar esta hipótese. De qualquer modo, decidimos incluir dados desta região no estudo levado a cabo.

Tanto a riqueza alélica, como a riqueza alélica privada, apresentarem valores consideravelmente mais elevados no Nordeste Africano e na Península Arábica. A variação obtida, a nível de cada um dos países analisados, apontou o Sudão e o Iémen como os países com os valores mais elevados de riqueza alélica. Cumulativamente o Iémen apresentava também os valores mais elevados de riqueza alélica privada. Estes resultados apoiam o Nordeste Africano como um putativo centro de origem do burro doméstico, no entanto, os níveis elevados de riqueza alélica única no Iémen, abrem a possibilidade desta região como um outro centro de origem do burro doméstico.

As ferramentas moleculares e as metodologias não-invasivas desenvolvidas para o estudo genético do burro selvagem Africano, provaram ser úteis no estudo de espécies aparentadas, tais como a zebra de Grevy. A zebra de Grevy partilha características ecológicas e comportamentais com o burro selvagem Africano e encontra-se atualmente distribuída na Etiópia e no Quénia, enfrentando um elevado nível de ameaça e estando categorizada como “ameaçada” (EN) pela IUCN.

O primeiro estudo genético não-invasivo das populações sobreviventes na Etiópia foi levado a cabo, usando a variação do ADN mitocondrial, para avaliar níveis de diversidade e estrutura populacional. Os resultados obtidos demostraram que as populações amostradas desta espécie apresentavam valores de diversidade extremamente baixos. As populações do Sul e do Centro da Etiópia não partilham haplótipos mitocondriais, indicando que o processo de isolamento entre estas populações tem vindo a decorrer há já algum tempo.

Summary

The African wild ass (*Equus africanus*) is one the eight extant equid species. This species carries the genetic inheritance of a 55-million-year evolutionary process, corresponding to the long evolutionary line of the Equidae family. However, it is unique among ass-like Equids, because it is the only species to have undergone a domestication process, resulting in the wide spread domesticated - the domestic donkey.

Historically, the African wild ass was distributed across a wide geographic range, north and west into Sudan and Egypt, however, and as result of human-mediated actions and environmental changes, its current distribution range is believed to be circumscribed to the Danakil Desert of Ethiopia and Eritrea, in a fragmentary distribution.

The great decline in population numbers, both in Eritrea and Ethiopia, has transformed the African wild ass into a vulnerable species, being currently considered the most endangered wild Equid by IUCN (International Union for Conservation of Nature) guidelines and facing an extremely high risk of extinction.

Given the current scenario it urges to develop and apply conservation measures which will prevent further depletion of diversity that compromises the survival of this species. But to

conserve, we needed first and foremost to assess relevant diversity and demographic parameters in extant populations in Eritrea and Ethiopia, and secondly to clarify the phylogenetic relationships between the African wild ass and the remaining ass-like species (*Equus hemionus and Equus kiang*), which remain very poorly known. Finally, due to its close genetic proximity to the domestic donkey, it is also important to assess the contribution of the African wild ass to the genetic pool of the domestic donkey and, thus, contribute to further clarify the complex domestication process of the donkey.

Reassessment of the evolutionary history of ass-like Equids was done by using a non-invasive sampling approach, collecting representative fecal samples from extant African and Asiatic wild ass populations across a vast area of their distribution range. Mitochondrial DNA (mtDNA) sequencing analyses was used to examine intraspecific genetic diversity and population structure, and to reconstruct phylogenetic relations among wild ass species/subspecies.

Analyses of obtained cytochrome b (Cyt *b*) sequences from extant African and Asiatic taxa showed that ass-like species shared a common ancestor approximately 2.3 Mya (95% CI; 1.4-3.2 Mya), what is somewhat coincident with the timeframe from the fossil record for the putative ancestor of the wild ass branch and thus supports the hypotheses of wild asses co-existing with early horses in North America.

Phylogeographic patterns among wild ass species were vastly different, however in what concerns the African wild ass, no geographic structuring of mtDNA haplotypes was obtained, with three out of four haplotypes been shared between Ethiopian and Eritrean populations. Obtained results, unraveled phylogenetic relationships among Asiatic wild ass species/subspecies that differ from currently accepted taxonomy and call for a more extensive reevaluation.

The first necessary step towards obtaining “neutral” genetic information on the African wild ass was taken by assessing levels of genetic diversity, population structure and demographic parameters, using nuclear microsatellite genotyping data.

Obtained diversity values clearly pointed to the population of Eritrea as the one possessing the highest diversity values. Analysis of migratory patterns allowed the detection of bilateral contemporary and historical migration between the Ethiopian and Eritrean populations. These findings are in line with the absence of geographic structuring in mtDNA haplotypes and once again confirmed by the absence of structure obtained by the analyses of nuclear data.

Previously reported mitochondrial DNA haplotypes observed in domestic animals were detected in individuals morphologically identified as African wild ass (*Equus africanus*), in both the Eritrean

and the Ethiopian populations. Despite limitations imposed by the reduced number of microsatellites, in the identification of putative hybrids between the wild and the domestic forms, it was possible to identify one individual in the Eritrean population that presented an admixture pattern consistent with a first generation hybrid (F1). Our results point towards the existence of sporadic and geographically limited hybridization between the wild and domestic forms.

Due to its close genetic proximity, the African wild ass is considered as the most probable ancestor of the domestic donkey. The identification of two clear lineages (named Clade 1 and Clade 2) among domestic donkey mtDNA control region haplotypes, is believed to be the result of two separate domestication events, from two ancestral wild populations. The existence of two morphological distinct subspecies of the African wild ass – the Nubian and the Somali wild ass – has also been documented. Previous work on donkey domestication has identified the Nubian wild ass (*Equus africanus africanus*) as the probable ancestor of Clade 1 donkeys.

Our analyses of control region haplotypes from ancient, historical and museum specimens of morphologically identified African wild asses, has confirmed the central role of the Nubian wild ass in Clade 1 ancestry. In fact, Nubian wild asses and domestic donkeys of Clade 1 are almost indistinguishable on the basis of mtDNA, with Nubian wild ass samples possessing haplotypes identical to those of domestic donkeys of Clade 1.

Clade 1 topology is consistent with a scenario whereby the Nubian wild ass was domesticated in several areas and/or over an extended long period, with multiple recruitments from the wild.

The extended mitochondrial dataset from free-living Somali wild ass showed that the Somali wild ass is distinct from Nubian wild ass and from domestic donkeys of both Clades 1 and 2. Additionally, the low variation and large sample size of Somali wild ass made it unlikely that additional lineages will be identified in the future and, thus, make Somali wild ass a less probable candidate for Clade 2 ancestry.

Presently, the archaeological findings point towards two regions as the centers of origin of the domestic donkey – Northeast Africa and Southern Arabian Peninsula. Here we have assessed and compared levels and patterns of genetic diversity, using microsatellite variation from eight donkey populations, representing the three hypothesized centers of origin of Northeast Africa, Near East and the Arabian Peninsula. Although the Near East has also been suggested as a center of origin, in the last decades archaeological data has dismissed the role of this region. Nonetheless we decided to include this region in our study.

Both allelic and private allelic richness presented considerable higher values in Northeast Africa and in the Arabian Peninsula. Variation at country level revealed Sudan and Yemen as the countries possessing higher allelic richness and cumulatively Yemen presented also higher values for private allelic richness. Obtained results support previously proposed Northeast Africa as a putative center of origin, but the high levels of unique diversity in Yemen opens the possibility of this region as yet another center of origin for the domestic donkey.

Developed molecular tools and non-invasive methodologies for the genetic study of the African wild ass, have proven to be useful for the study of closely related species such as the Grevyi´s zebra.

Grevyi´s zebra share ecological and behavioral features with the African wild ass and are currently distributed in Ethiopia and Kenya, facing a high level of threat and being categorized as endangered by IUCN.

The first non-invasive genetic study on extant Grevy´s zebra populations in Ethiopia was conducted, using mtDNA variation in order to assess levels of diversity and population structure. Results showed that populations of this species presented extremely low diversity values. Southern and central Ethiopian populations did not share haplotypes, indicating that the process of isolation between these populations have been undergoing for a long period of time.

Palavras-chave / Keywords

Equid / Equídeo

Evolutionary history / História Evolutiva

African wild ass / Burro selvagem Africano

Asiatic wild ass / Burro selvagem Asiático

Conservation genetics / Genética da Conservação

Genetic diversity / Diversidade Genética

Phylogeography / Filogeografia

Non-invasive sampling / Amostragem não-invasiva

Mitochondrial DNA / ADN mitocondrial

Microsatellite / Microsatélites

Domestication / Domesticação

Domestic donkey / Burro doméstico

Index of contents

Chapter I: General Introduction

Chapter II: Evolutionary history of ass-like equids

Chapter III: Conservation genetics of the critically endangered African wild ass (*Equus africanus*): a non-invasive molecular approach

Chapter IV: Insights into donkey domestication and ancestry: patterns of past and current variation

Chapter V: Ancillary work

Index of Tables

Chapter II: Evolutionary history of ass-like equids

Chapter III: Conservation genetics of the critically endangered African wild ass (Equus africanus): a non-invasive molecular approach

Chapter V: Ancillary work

Index of Figures

Chapter I: General Introduction

Chapter II: Evolutionary history of ass-like equids

Abbreviations

AD – Anno Domini

aDNA – ancient DNA

AIC – Akaike information criterion

AMOVA – Analyses of molecular variance

BCE – Before the Common Era

BP – Before Present

bp – Base pairs

BSP – Bayesian skyline plot

CI – Confidence interval

CIBIO – Research Center in Biodiversity and Genetic Resources

CITES - Convention on International Trade in Endangered Species of Wild Fauna and Flora

CR – Control region

Cyt *b* – Cytochrome b

DNA - Deoxyribonucleic acid

ESS – Effective sample size

ESU – Evolutionary significant unit

EWCO - Ethiopian Wildlife Conservation Organization

FCA - Factorial Correspondence Analysis

FCT - Portuguese Foundation for Science and Technology

GTR – Generalized time reversible

HKY – Hasegawa, Kishino and Yano substitution model

HVRI – Hypervariable region 1

IAM – Infinite allele model

IUCN – International Union for Conservation of Nature

LD – Linkage disequilibrium

LGM – Last Glacial Maximum

MCMC – Markov chain Monte Carlo

mtDNA – Mitochondrial DNA

Mya – Million years ago

PCoA – Principal coordinates analysis

PCR – polymerase chain reaction

PIC – Polymorphic information content

rRNA - Ribosomal ribonucleic acid

SNP – Single nucleotide polymorphism

TMRCA – Time to the most recent common ancestor

TPM – Two-phase mutation model

Chapter I

General Introduction

1.1 Evolutionary history of Equids

Equids (i.e. Equidae family) including the extant horses, asses, zebras and many other extinct horse-like species have been the object of interest from natural historians, since the early stages of history. Much of this attention can be attributed to the long time association between humans and horses, with the oldest records emerging in Paleolithic cave painting in France (Les Combarelles) and Spain (Altamira), between 35000 and 100000 years ago (MacFadden 1992).

The ancient Greek philosopher and naturalist Aristotle (384-322 BCE) was among the firsts to observe and register the existence of different kinds of "horses", stating: "*there is a genus of animals that have manes, as the horse, the ass, the oreus, the ginnus, the innus and the animal which in Syria is called heminus (mule)*" (MacFadden 1992).

Only much later, as modern Biology arose, by early eighteenth century did the name "*Equus"* became recognized as one of the thirty-three mammal genera included in Carl Linnaeus's first edition of *Systema Naturae*.

The first horse-like fossil bones ever analyzed were dug from the Montmartre gypsum in the city of Paris, France, in early ninetieth century. Latter, those bones were taken to the Paris Conservatory and analyzed by the notorious scientist Baron Georges Cuvier, who in 1825 illustrated and described the remains as a tapir-like browser equid - the *Paleotherium*.

The next breakthrough in the study of equid fossils was achieved in 1839 by Sir Richard Owen, who descibed and named the fossil equid *Hyracotherium* from the early Eocene London Clay.

By 1854, the North american scientist Joseph Leidy, described the first New World fossil horse genus, from Noth America, as *Hippodon* and latter, in 1857, another form was yet descibed by him, the *Merychippus*.

Despite this apparently isolated findings, the area of fossil horse systematics begun to gain momentum, with groups of paleontologists across the Atlantic Ocean working in the analyses of separate collections. In a social and scientific context where old (Creationism) and new (Evolutionism) paradigms were at the center of debate, the equid fossil register became unexpectedly the most powerful weapon in favor of the Evolutionary theory. Thomas Huxley, a notorious evolutionist, visited the United States of America in 1876 on a lecture tour to advocate for the Evolutionary theory. Before his talk in the New York Academy of Sciences, he spent some time in Yale, studying the fossil horse collection assembled by the paleontologist Othaniel Charles Marsh during expeditions in the western territories. Huxley was so overwhelmed by the richness

of this assembly that he realized it stood as the strongest evidence ever presented in favor of Evolution (MacFadden 2005). From that point in time on, transatlantic scientific awareness and communication improved and led to both schools accepting the idea that the main adaptive radiation of post-early Eocene fossil horses occurred in North America, with parallel evolution taking place in Europe.

Despite the remarkable improvement in the knowledge of equid evolution, most evolutionists saw the equid phylogeny as linear (orthogenetic), with species gradually evolving into the next (and therefore "higher") form (Bennett 1992). This simplistic approach was only rejected in the beginning of the twentieth century, with paleontologists understanding that the pattern of Equid evolution would better be represented by a tree with numerous "side branches", some leading to species already extinct and others leading to species closely related to extant *Equus* species (Fig. 1) (MacFadden 2005).

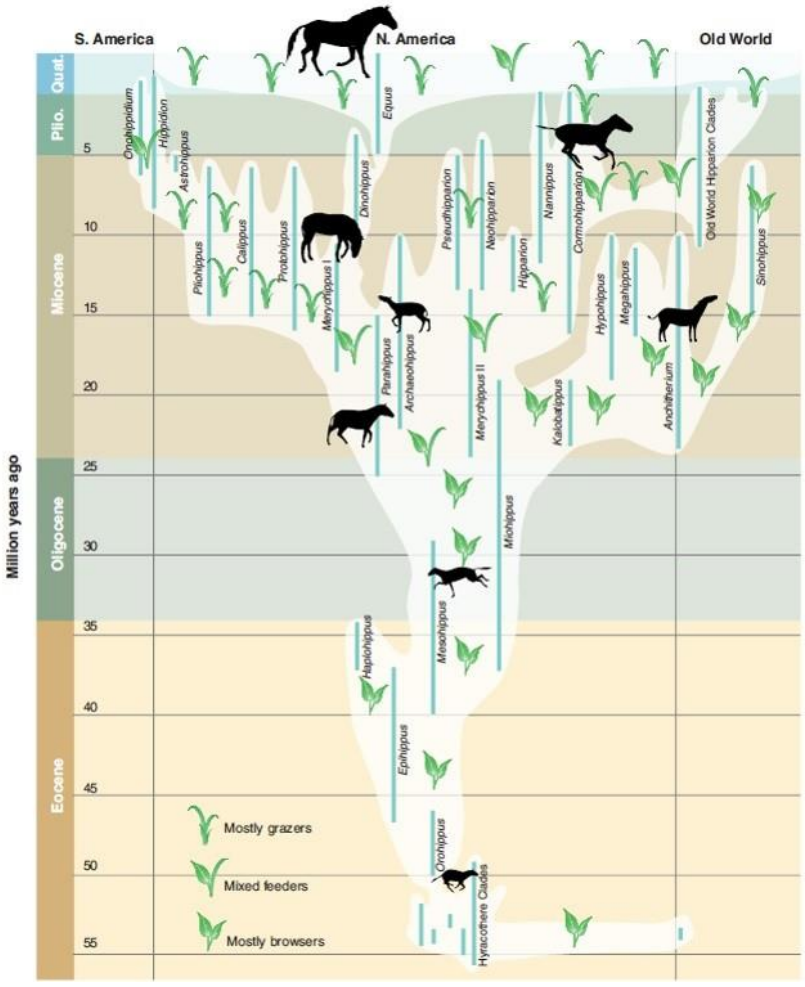


Figure 1 - Phylogeny, geographic distribution, diet and body sizes of the Equidae family over the past 55 million years (MacFadden, 2005).

The starting point to taxonomically place equids among mammals done by Linnaeus in 1735, was later revised in 1766, in the final version of *Systema Naturae*. In this revision, he grouped *Equus*, along with hippopotamus and tapirs (as a species of the later) in a taxonomic group named Belluae, within Ungulata. This classification did not recognize Perissodactyla as a natural group. In 1816, de Blainville classified soliped horses (i.e. *Equus*) and pachyderms within the "Impairs" – forms with non-paired digits - however it was Owen (MacFadden 1992) to take credit for both the naming and the concept of Perissodactyla (odd-toed) and Artiodactyla (even-toed). Owen's classification viewed Perissodactyla as a monophyletic clade, incorporating both extant and fossil equids. Different classifications arose in time due to the analyses of different morphological characters. Although the best systematic methodology would incorporate as many different characters as possible, the most common morphological features used in the phylogenetic assessment in equids were dental, cranial and postcranial characters.

Despite the many updates and revisions in equid systematics through time, the general pattern of the Equidae phylogeny has remained relatively unchanged since the middle of last century, with emphasis being placed on vertical classification and recognition of strictly monophyletic clades. It is currently accepted that Equidae had its origin in the Eocene, approximately 55 million years ago, being represented by three dozen extinct genera and a few hundred extinct species (MacFadden 1992). This once greatly diverse group of animals is now represented by the single modern genus, *Equus*.

1.2 The *Equus* genus

1.2.1 *Equus* taxonomy

Equus is the only remaining genus of the Equidae family, incorporating both extant and fossil species. The taxonomic literature on *Equus* contains approximately 230 named species (Winans 1989), many of these identified on the basis of fragmentary material and resulting in inadequate descriptions. Currently, *Equus* is represented by eight extant species: domestic horse (*E. caballus*), Przewalski´s horse (*E. przewalskii*), kiang (*E. kiang*), Asiatic wild ass (*E. hemionus*), African wild ass (*E. africanus*), mountain zebra (*E. zebra*), plains zebra (*E. quagga*) and Grevy´s zebra (*E. grevyi*) (Moehlman 2002). Under this taxonomy, domestic donkey (*E. africanus asinus*) is considered as a subspecies of African wild ass.

Many morphological features have been used in order to group individuals belonging to this genus, however few studies managed to integrate all the information to produce a consensual

phylogeny. Up till the emergence of modern molecular techniques, such as DNA sequencing, the most widely used characters were morphological such as skull, limb and teeth features. In particular the morphology of the metaconid-metastylid loops of the double knot and the shape of the entoflexid (linguaflexid) separating them has been used to separate two main groups: the stenoids (zebras and asses), with v-shaped entoflexids and the caballoids (true horses), with u-shaped entoflexids. As molecular biology techniques quickly evolved, a new array of potential characters became available for discerning the systematics of extant taxa.

1.2.2 *Equus* early history and dispersal patterns

According to Lindsay (1980), the *Equus* genus was first represented in the fossil record by *Equus simplicidens* (Cope), occurring for the first time in the Hagerman fauna, dated at about 3.3 Ma (million years) (Kurtén 1980). Excavations in the Ringold Formation in Washington, where the same species was identified were dated even later at 3.4-3.8 Ma (Lindsay 1984). In the Concha Fauna, New Mexico, a fossil horse described as an early *Equus* or the direct ancestor of *Equus* (an advanced *Pliohippus*) was identified and dated at approximately 4.2 Ma, what is in good agreement with the 3.9 Ma age for the common ancestor of all extant *Equus* proposed in a study by George and Ryder (1986). Early *Equus* were identified as belonging to the stenoid group on the basis of tooth morphology, however diversification in morphological characters (i.e. limb proportion, body size) took place prior to the first dispersal events to Eurasia, believed to have occurred 2.5-3.0 Ma ago (Lindsay 1980; Azzaroli 1982). The first forms identified in the Old World were closely related to *E. simplicidens* in terms of dental features, however as in the New World, *Equus* became morphologically and taxonomically diverse, with simpatry occurring mostly between two species (Forstén 1988).

Zebras species have their probable origin in early stenoids, arriving first to Eurasia and dispersing latter to Africa, where they were first identified in a site in the Omo beds (Ethiopia) at about 2.0 Ma (Hooijer 1976; Eisenmann 1976). Two different species were identified at this site, *Equus quagga* and *Equus grevyi* (or an ancestral of this species) (Eisenmann 1985). *Equus zebra* only shows up in the fossil record by later middle Pleistocene (Churcher 1978).

The fossil record of asses is ambiguous. African and Asiatic asses are, in terms of karyotype, more closely related to each other than to other stenoids (Ryder *et al.* 1978), however, they are dissimilar enough in order to justify long, separate evolutionary histories. *Equus cumminsi* (Cope), allegedly from the middle Blancan (3.2-4.2 Ma) and early Pleistocene (<1.8 Ma ago) of Kansas

and Texas, has been identified as an early ass on the basis of dental morphology. Dalquest (1978) argued that ass-like equids were already present in North America by the early Pleistocene, at about 1.8 Ma ago. The first true African ass was described by Churcher (1982), from limited data collected in Olduvai Gorge, Tanzania. This fossil remains were dated at about 1.5-1.7 Ma. Ancestry and dispersal of extant ass-like equids, in the Old world, is still unclear. The absence of the African wild ass in Asia's fossil record makes the dispersal patterns difficult to track. It is believed that the Indian subcontinent might have worked as a possible evolutionary and dispersal center of the asses, from an early immigrant stenoid/asslike ancestor (Forstén 1992).

True horses or caballoids are believed to have appeared for the first time in North America. Dates attributed to early caballoids from North America (approximately 1.8-0.3 Ma) are older than those of their early occurrence in the Eurasian fossil record (1-0.4 Ma ago). Caballoids became common toward the middle or the end of that period, replacing almost totally the stenoid lineages (Forstén 1992).

1.2.3 Molecular studies on the *Equus* genus

Early molecular studies on extant equids, focused on the analyses of banded karyotypes, demonstrating the existence of extensive karyotipic rearrangements since divergence from a common ancestor (Ryder *et al.* 1978). Subsequent studies on equid hemoglobins (Ryder *et al.* 1979) provided little improvement in the clarification of phylogenetic questions, however a study by Kaminski (1979), on equid serum esterases, provided some insights into phylogenetic relationships among equid species. The first study to use mitochondrial DNA, (restriction-endonuclease maps) in order to compare equid species, was published George and Ryder (1986). This study was also the first to calculate nucleotide sequence divergence among this species and to date the divergence time of extant lineages from a common ancestor at about 3.9 Million years before present. Main findings in these study supported the existence of tree major clades; one grouping all zebras (monophyletic), a second grouping *E. africanus* and *E. hemionus* and a third that associated the true horses (*E. przewalskii* and *E. caballus*). Discrimination of the branching order of *E. africanus* and *E. hemionus* failed to be achieved.

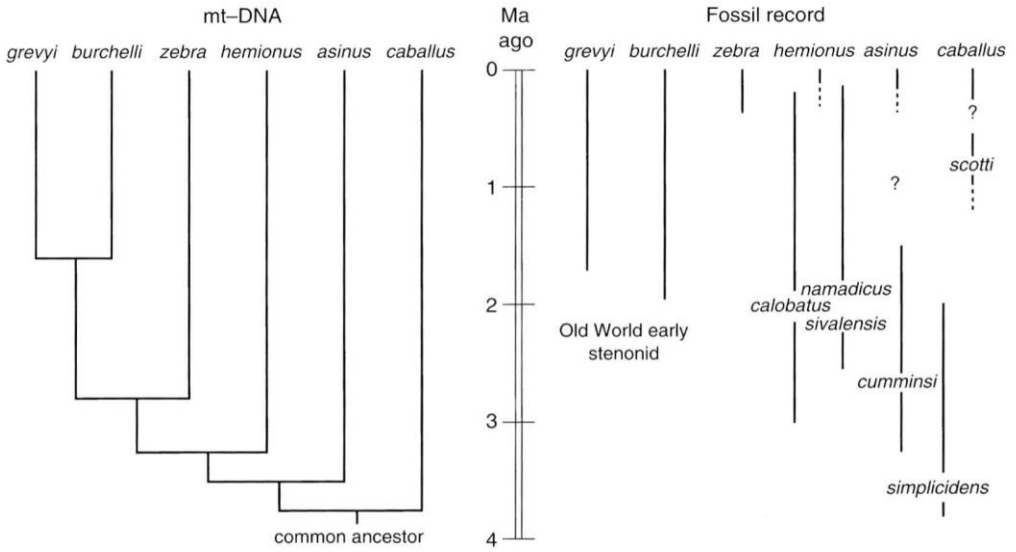


Figure 2 - Phylogenetic trees inferred from mtDNA cleavage maps (George & Ryder 1986) and the fossil record (Forstén 1992).

Subsequent molecular studies using globin genes (Oakenfull & Clegg 1998) showed an early divergence of the horse and the zebra/ass lineages (Fig. 2), with branching orders within the zebra/ass clade presenting low bootstrap support values.

Shortly after, Oakenfull *et al.* (2000) performed a mitochondrial DNA survey among extant equid species, using samples from captive individuals. This study included the phylogenetic analyses of two mitochondrial DNA genes (control region and 12S rRNA) for approximately thirty individuals, including several subspecies. Obtained results showed that the interspecies branching with the highest bootstrap support was that separating horses (*E. caballus* and *E. przewalskii*) from other equids. With the exception of the Asiatic asses (*E. hemionus* and *E. kiang*), individuals belonging to the same species grouped together with high bootstrap values. Branching order of the zebra and ass species was different depending on the methodological approach (maximum likelihood, maximum parsimony and neighbor-joining). Dating of the common ancestor of extant equids was estimated to be 2.3 Ma ago, what is in agreement with pervious results using globin genes, that pointed a date around 2.4 Ma ago, but in clear disagreement with both Ryder's study (1978) and the fossil record pointing towards an approximate date of 3-4 Ma ago for the common ancestor of *Equus*. Oakenfull bridged these differences by defending that the common ancestor of extant equids was a species that existed around 2 Ma ago, but was not one of the first *Equus* to have emerged in the New World, approximately 3.7 Ma ago.

More recently a study on the evolutionary history of Peryssodactyls (Steiner & Ryder 2011), using both mitochondrial and nuclear markers and a modern Bayesian approach confirmed that the deepest divergence of the phylogeny corresponded to the split between caballin and noncaballine equids. Within noncaballines, the African wild ass was identified as the sister taxon of two monophyletic clades composed of the Asiatic asses (*E. kiang* and *E. hemionus*) and zebras, with the mountain zebra (*Equus zebra*) having diverged first relative to Burchell's (*E. quagga burchelli*) and Grevy's (*E. grevyi*) zebras. *Equus* was estimated to have emerged during the Pliocene (3.8 Mya). Estimated divergence of African wild ass from its common ancestor with caballine horses was dated at approximately 2.1 Mya, which is slightly prior to the diversification of Asiatic asses and zebras at about 1.9 Mya. Asiatic asses were estimated to diversify more recently in comparison to extant zebras, with the kiang around 500 000 years ago and Asiatic wild asses (onager and kulan) at about 300 000 years ago.

Another recent study, by Steiner *et al.* (2012) used 22 nuclear and mitochondrial genes in order to infer the molecular phylogeny of extant equids and clarifying the effect of ancestral polymorphism in the resolution of this questions. Despite limited information from low variation at nuclear loci, same further insights were obtained, namely the monophyletic origin of all ass species, with the African wild ass as the sister taxon of Asiatic asses (combined data set). Relationships among zebras were well supported in a monophyletic clade, placing the mountain zebra as a sister taxon to Burchell's and Grevy's zebra. The lowest bootstrap support values corresponded to the dichotomy between asses and zebras, and the clade comprising *E. hemionus* subspecies.

One of the most recent study in *Equus* evolution, was published by Orlando *et al.* (2013), and used genomic data in order to recalibrate divergence times among extant species. This study proposes a 4.0-4.5 Myr BP for the most common recent ancestor of all living Equus.

1.3 The wild-ass branch of the *Equus* phylogeny

The wild ass branch of the *Equus* phylogeny includes three taxonomically distinct species: the African wild ass (*Equus africanus*), the Asiatic wild ass (*Equus hemionus*) and the Tibetan wild ass (*Equus kiang*).

Phylogenetic assessment of species' relationship whit in the ass-like equids group of animals, has been a difficult task, even with the improvement in both available molecular markers and phylogenetic inference methodologies. Mostly because all studies aiming to reconstruct the

phylogeny of extant equids, in particular the wild-ass branch, have lacked enough information from individuals belonging to extant natural populations, as only captive animals, which often have unclear origins. This might be a consequence of their geographical locations in extreme environments (Horn of Africa, central Asia Soviet Union republics, Iran and the Tibeten Plateau), u, making access to samples from these populations extremely difficult.

1.3.1 The Asiatic wild ass - *Equus hemionus* (Pallas 1975)

There are five recognized Asiatic wild ass (*Equus hemionus*) subspecies (Grubb 2005): *Equus hemionus hemionus* (Mongolia), *Equus hemionus khur* (India), *Equus hemionus kulan* (Turkmenistan, re-introduced in Kazakhstan, Uzbekistan and Ukraine), *Equus hemionus onager* (Iran) and *Equus hemionus hemippus* (extinct, formerly from Syria south into the Arabian Peninsula). *Equus hemionus luteus* is most probably a synonym of *Equus hemionus hemionus* (Oakenfull *et al.* 2000; Grubb 2005).

Historically the Asiatic Wild Ass ranged through much of Mongolia, north into Russia, east to north-eastern Inner Mongolia and Manchuria (China), and west to Dzhungarian Gate (Grubb 2005). It formerly occurred in Kazakhstan, Anatolia (Turkey), Syria, Iran, northern Iraq, Afghanistan, Pakistan and India (Grubb 2005). It also extended through the Arabian Peninsula as far south as central Saudi Arabia. It survived in Armenia and Azerbaijan until the 17th-18th centuries and became extinct Syria, in 1927 (Moehlman *et al.* 2008a).

Currently the most abundant subpopulation of the species occurs in the southern part of Mongolia and adjacent northern China (Feh 2002). The species also survives as isolated populations in the Rann of Kutch (India), the Badkhyz Preserve (Turkmenistan), Touran National Park and in the Bahram Gour Reserve (Iran) (Feh 2002).

During the past century the total population and distribution range of the Asiatic Wild Ass has suffered a significant decline (Clark & Duncan 1992; Feh 2002). Southern Mongolia is the home of the largest population of Asiatic Wild Ass, representing almost 80% of the global population (Feh 2002), despite having suffered a significant decline from approximately 40000 individuals in 1997 to an estimated 18000 in 2003. The second largest population is the Indian Khur (*E. h. khur*) with an estimated population in 2004 of 3,900 in the Little Rann of Kutch (Moehlman *et al.* 2008a).

The Kulan (*E. h. kulan*) populations in 2005 were approximately 1,300 in Turkmenistan (Badkhys Reserve and seven reintroduction sites (Moehlman *et al.* 2008a). Numbers of the Iranian Onager

(*E. h. onager*) are roughly estimated at 600 individuals in the two protected areas (Moehlman *et al.* 2008a).

Overall population trend shows a decreasing tendency, with extant populations being placed in different IUCN categories according to the level of endangerment. *E. h. kulan and E. h. onager* are Critically Endangered (CR), *E. h. khur* is Endangered (EN) and *E. h. hemionus* is Vulnerable (VU).

Major threats to extant populations of Asiatic Wild Ass are habitat loss as a result of human settlements, competition for water and resources with domestic livestock, poaching and stress events (such as droughts or diseases). Besides this more direct causes, as Asiatic wild ass populations became smaller and more isolated, they also became more demographically and genetically vulnerable and hence facing a higher risk of extinction. Asiatic Wild Ass inhabit mountain steppe, steppe, semi-desert and desert plains. The species present a feeding strategy similar to that observed in other equids in xeric environments, being predominately grazers, when grass in abundant and browsing a large portion of their diet in the dry season or in drier habitats (Moehlman *et al.* 2008a). Breeding is seasonal and females with young foals tend to group in small groups, of two to five females (Moehlman *et al.* 2008a). Male breeding strategies differ considerably, with studies identifying two main strategies: individual stallions either defend territories or form all-male groups (Feh 2002).

1.3.2 The Tibetan wild ass - *Equus kiang* (Moorcroft 1841)

Equus kiang distribution range is centred on the Tibetan Plateau between 2,700 and 5,400 meters. Most of the distribution is in China, however the range extends into northern parts of Pakistan, India, Nepal and possibly Bhutan. The current global population estimate of Kiang is 60,000-70,000 animals, 90% of which are in China, in the provinces of Qinghai, Gansu, Xinjiang, and Tibet (Xizang) (Shah 2002).

It is listed as Least Concern (LC) by IUCN, due to its wide distribution, large population, and because a fast decline of extant populations is not currently expectable (Shah 2008) .

Its main threats to this species are potential conflicts with humans and their livestock, over-hunting, and possible disease transmission from domestic livestock.

Equus kiang is an animal of open landscapes, inhabiting plains, valleys and hills, wherever suitable forage is abundant. It has a social organization similar to other wild equids such as the

African and the Asiatic wild ass, with solitary males occupying vast areas males at least during part of the year. Females in similar reproductive stage can often associate in groups, and large groups of individuals can form whenever good pasture is available (usually during fall and winter) (Schaller 1998). The kiang has its peak breeding and foaling from June to September and mares have a gestation period of 355 days.

There has been considerable debate about the taxonomy of *Equus kiang*. Up until late last century, the kiang was considered to be a subspecies of *Equus hemionus.* The species status was revised by Groves (1967) and later by Bennett (1980), who agreed in separating *kiang* from *hemionus*. Three subspecies of *Equus kiang* were recognized, on the basis of morphological features such as skull proportions, angle of incisors, shape of rump, coat colour, and body size (Groves 1967): *E. k. kiang*, *E. k. holdereri*, and *E. k. polyodon*. These subspecies are known by their geographic range as the Western Kiang, Eastern Kiang, and Southern Kiang, respectively. Validity of this subspecies is questionable (Shah 2002), though this has not been officially documented.

1.4 The African wild ass – *Equus africanus* (von Heuglin & Fitzinger, 1866)

1.4.1 Nomenclature, morphological features and distribution ranges

Historically there were three recognized subspecies: the Atlas wild ass - *Equus africanus atlanticus*; the Nubian wild ass (Heuglin and Fitzinger, 1866) – *Equus africanus africanus* and the Somali wild ass (Noack, 1884) – *Equus africanus somaliensis* (Fig. 3).

Figure 3 - Atlas wild ass (A), Nubian wild ass (B) and modern Somali wild ass (C) (Kimura et *al*, 2010).

The Atlas wild ass has been recorded in prehistoric rock art in Algeria, from which it was possible to observe the presence of a well-developed shoulder stripe and striped legs (Fig. 3A). Unlike the Atlas wild ass and the Somali wild ass (Fig. 3C), the Nubian wild ass (Fig. 3B) possessed no leg stripping. Somali wild asses is distinguishable by the clear banding around the legs, while it may or may not have a shoulder stripe. Somali wild asses are long-legged, powerful equids, presenting a short and smooth coat, of a light grey to fawn colour, which fade to white on the undersides and legs (Clutton-Brock 1992).

The Atlas wild ass had its distribution range from the Atlas region of north-western Algeria, to adjacent areas of Morocco and Tunisia, however its believed to have become extinct at about 300 AD as a result of sports hunting by ancient Romans (Van Bemmel 1972). The Nubian wild ass inhabited the Nubian desert of northeastern Sudan, from east of the Nile River to the shores of the Red Sea and south to the Atbara River and into northern Eritrea. This subspecies is now extremely rare or even extinct in the wild. The Somali wild ass, was found in the Eritrean region of Denkelia, in the Danakil Desert and the Awash River Valley in the Afar region of north-eastern Ethiopia, western Djibouti and into the eastern Ethiopian region of Ogaden. In Somalia, they ranged from Meit and Erigavo in the north to the Nugaal Valley and as far south as the Shebele River (Moehlman 2002). The Somali wild ass still subsists in regions in Eritrea and Ethiopia, however its presence in Somalia is currently uncertain.

1.4.2 Population estimates and trends

In Ethiopia, there has been a severe population decline since the early 1970s (approximately 95%). The last comprehensive survey across the historic range of the African wild ass in Ethiopia took place in 2007 (Moehlman *et al.* 2008b). Obtained data revealed that this species have been extirpated from the Yangudi-Rassa National Park and the Somali Region and that the only remaining population lives now in the northeastern Afar Region. The total number of individuals observed during this survey was 25 in an area of 4,000 km² yielding a density of 0.625 animals per 100 km². The population in the Serdo-Hillu area has remained stable over the last 10 years, mainly due to a conservation program and the combined efforts of the Ethiopian Wildlife Conservation Authority and local Afar pastoralists. Despite the undergoing protective measures, this population remains under high risk of extinction. Rough estimates for the number of individuals, point towards as few as 200 African Wild Asses left in Ethiopia. Long-term data are not available in Eritrea, but since the mid-1990's the population appears to be stable and in a limited study, in the Northern Red Sea zone, the density is approximately 47

individuals per 100 square kilometers (Moehlman 2002; Moehlman *et al.* 2008b). Recent research indicates that African Wild Ass currently inhabit approximately 11,000 km² in the Denkeli desert (Teclai 2006), however surveys are needed to determine the distribution and density of African Wild Ass in this larger area. Roughly 400 African wild asses are estimated to subsist in Eritrea.

It is unknown if African wild ass currently persist in Somalia. Some animals may remain near Meit and Erigavo, however the most numerous population in northern Somalia (from the Nugaal Valley to the Djibouti border), with a past estimated population of 4,000–6,000 individuals, had practically vanished by 1997.

Overall number of African wild asses in the wild is estimated in less than 600 animals and the species is currently listed as critically endangered by IUCN (Moehlman *et al.* 2008b) and listed in CITES Appendix I.

1.4.3 Habitat and Ecology

The African wild ass inhabits areas of arid and semi-arid bushland and grassland. They are able to subsist in extreme environments such as the volcanic landscape of the Great Rift Valley, ranging from below sea level (Dalool Depression) to an altitude of approximately 2000 meters. Limited observations in Eritrea and Ethiopia, indicate that African wild asses are primarily grazers (Kebede 1999), however they can also browse.

The African wild ass exhibits the social organization typical of equids that live in arid habitats (Moehlman 1998). Groups may include up to five individuals, however the only stable groups are composed of a female and her offspring. The structure of temporary groups might vary from single-sex adult groups to mixed groups of males and females of all ages. Adult males are frequently territorial and solitary, but can also associate with other males. Low sociability, in comparison to other equid species, may be due to low forage quality and availability, which does not allow females to forage in close proximity for long periods of time.

The African wild ass has a resource-defense polygyny mating system in which males defend mating territories that contain the resources (water and forage) required by females (Moehlman 2002). Limited data indicates that females have their first foal at age three to four years and will typically have a surviving foal every other year. Lactating females are water stressed, drinking three times as often as other adults in the population and tend to stay within one kilometer of water what makes them and their foals at higher risk of predation.

1.4.4 Major threats, legal protection and conservations actions

The main threats to African wild ass survival are illegal hunting for food and medicinal purposes, competition with livestock for vegetation and water and the possible interbreeding with the domestic donkey (Moehlman 2002).

In Ethiopia and Somalia local pastoralists kill African wild ass for food and/or medicine. The lack of proper medical assistance in these regions makes the use of traditional medicine a common practice. Body parts and soup made from bones are used for treating tuberculosis, constipation, rheumatism, backache, and bone ache (Kebede 1999; Moehlman 2002). In Eritrea, the Afar pastoralists do not shoot wildlife and in that way humans do no stand as direct threat, *per se*.

Limited access to drinking water and forage (largely due to competition with livestock) is a major constraint in this extremely depleted environment, particular in years of drought. The third major threat to the survival of the African Wild Ass is possible interbreeding with the domestic donkey (Moehlman 2002).

In Ethiopia, the African wild ass in protected by wildlife laws, meaning that it cannot be hunted or killed, with no exceptions or special permits. Two protected areas (Yangudi-Rassa National Park and Mille-Serdo Wild ass Reserve) have been established, however this areas continue to be utilized by pastoralists and their livestock, hence Ethiopian Wildlife Conservation Organization (EWCO) possess insufficient funds or personnel for their appropriate management (Kebede 1999). In Eritrea and Somalia, the African wild ass has no formal legal protection.

Given the level of endangerment undergone by this species, some actions are necessary to ensure their long time survival and conservation. In particular, improving the protection and management of existing populations, clarifying the genetic status of the two subspecies of African wild ass and the extent of interbreeding with the domestic donkey, extending surveys and improving monitoring of known populations and conducting research on basic biology, seasonal movements and interactions with livestock.

1.5 Conservation genetics of endangered species

1.5.1 Conservation Biology and Conservation Genetics

The discipline of Conservation Biology arises as a reaction of society in general and of academics in particular, to the unprecedented loss of diversity in biological systems across the world. The term “Conservation Biology” was initially introduced as a conference title, held at the University of

California, in 1978, organized by biologists Bruce Wilcox and Michael Soulé. The main outcome of this meeting was the realization that this area of study would ultimately have as an objective to bridge the gap existing between theory in ecology and population biology and conservation policy and practice. Prior to this meeting, in 1974, Otto Frankel had already published a landmark work, "Genetic conservation: our evolutionary responsibility", that highlighted the importance of setting out conservation priorities from a genetic perspective, stating "*Genetic wildlife conservation makes sense only in terms of an evolutionary time scale. Its sights must reach into the distant future".*

Conservation Biology is without a dought an ambitious discipline. It faces complex problems with complex solutions, envolving not only scientific questions but also social and politycal actions. Conservation actions need to happen in the present, in order to guarantee the sustainability of ecosystems in the future, however, the present is dominated by the immediate economic interest of human populations. This balance, between what needs to be done in order to conserve species and ecosystems and what realistically can be done in order to maintain or increase social wellfare of human populations, is the key for the long term survival of species.

The applications of genetics to conservation had its foundation with the publications by Soulé and Wilcox (1980) and Frankel (1981). Conservation genetics aims to apply genetic methods to the conservation and restoration of biodiversity. It encompasses genetic management of small populations, resolution of taxonomic uncertainties, definition of management units and forensic analyses in order to understand species' biology (Frankham 2003).

1.5.2 Conservation genetics of small and endangered populations

It is currently accepted that both deterministic (habitat loss, over exploitation, introduced species and pollution) and stochastic (demographic, environmental, genetic and catastrophic) factors have a combined effect in biological extinctions (Shaffer 1981; Frankham 2005).

Deterministic factors are caused by humans and frequently related to human population growth. These factors are responsible for the decrease in population sizes, making them more susceptible to stochasticity and catastrophes (Frankham 2005).

Many species with once large population sizes and continuous distributions are now fragmented, subsisting in isolation, due to human mediated habitat loss.

Genetic diversity is a particular important genetic parameter to be evaluated, because it has long been seen as the raw material for evolution, on the face of environmental change (Allendorf 2007). The amount of genetic variation within a population is a good indicator for predicting its evolutionary potential. A decrease in genetic variability could have devastating consequences to populations, leading to an increase in the levels of inbreeding and consequently a reduction in fitness. Genetic diversity is lost in small mating populations at the same time they become inbred, so the two processes are closely related (Frankham 2005).

Inbreeding, the result of matting between related individuals, has long been known to reduce reproduction and survival in naturally outbreeding species (inbreeding depression). Inbreeding has deleterious consequences on all aspects of reproduction and survival, including sperm production, mating ability, female fecundity, juvenile survival, mothering ability, age at sexual maturity and adult survival in animals (Frankham 2002). The effects of inbreeding depression and loss of genetic diversity can interact with demographic, environmental and catastrophic factors in an “extinction vortex” (Frankham 2005).

Measuring the size of a population is also a central concept in conservation genetics of endangered populations. Population size can be standardized to its effective population size (*Ne*), which is defined as the number of individuals that would give rise to the calculated inbreed coefficient, loss of heterozygosity or variance in allele frequency if the population behaved as an idealized one (Frankham 2005).

Evolutionary processes in small and endangered populations are vastly different in small populations that in large populations. In small and endangered populations the role of chance (genetic drift) is enhanced, contrary to the effects of selection, which are vastly reduced or even eliminated.

1.5.3 Non-invasive genetic sampling

Assessing genetic parameters in wild populations was for a long time a complex task, which often implicated the killing of animals in order to perform scientific studies. Besides ethically questionable, destructive sampling was not possible in most endangered species, that are usually protected by international laws and directives. With the advent of polymerase chain reaction (PCR) and its wide use, it became possible to amplify DNA from small amounts of fresh, alcohol-preserved or even dry tissues (Taberlet & Luikart 1999). Non-invasive sampling techniques are

of particular interest for conservation biologists, because animals don't have to be captured, disturbed, or even observed, making genetic sampling much easier.

After an initial period of enthusiasm by the scientific community, some studies revealed a risk of genotyping errors associated with the use of hair or feces [see review by Beja-Pereira *et al.* (2009)]. Detecting and quantifying genotyping errors became a major concern when working in a non-invasive framework, and subsequent studies focused in improving the knowledge of which factors can influence genotyping quality (sampling preservation, DNA extraction, PCR improvement) as well as developing statistical methods to improve error detection.

Main errors associated with non-invasive sampling are allelic dropout (detection of a single allele of a heterozygous individual, usually due to sampling stochasticity), null alleles, false alleles (artefacts generated during the amplification process, misinterpreted as true alleles) (Taberlet & Luikart 1999). Such errors are the result of the frequently low quality/quantity extracts obtained by non-invasive samples, such as feces, hair, feathers.

Methodological approaches such as the "multiple tubes approach" (Taberlet *et al.* 1996), have vastly improved genotyping reliability in non-invasive samples, however a global approach that tracks and assess genotyping errors, from sample collection to genotyping results (Bonin *et al.* 2004) seems to provide a better control over the entire process, making monitoring of errors and contamination more efficient.

1.5.4 Molecular markers in Conservation genetics

The choice of the appropriate molecular marker will ultimately depend on the questions in study. The two most commonly used markers in non-invasive genetics are mitochondrial DNA and nuclear microsatellites, however new genomic approaches are becoming more and more common. Mitochondrial DNA is a maternally inherited circular molecule, containing in animals 36 or 37 genes and a non-coding sequence, termed the control region (CR). Vertebrate CR is commonly subdivided into three domains that differ from each other in base composition as well as in rate of evolution. The flanking domains (domain I and II) are hypervariable in base substitutions and indels, making these particular regions appropriate for intraspecific studies. Mitochondrial protein-coding genes such as cytochrome *b* are powerful markers for inferring evolutionary history at the families, genera and species level and in resolving taxonomic uncertainties in conservation genetics (Wan *et al.* 2004).

Microsatellites are stretches of short DNA sequences in which a motif of one to six bases is tandemly repeated. They are dispersed throughout the eukaryotic nuclear genome and inherited in a co-dominant Mendelian manner. Variability of this markers is frequently high, and in that way even a small number of loci, is usually enough to discriminate individuals. Besides being used in individual identification and in the inference of relationships, its neutral character makes them appropriate to assess population structure and demographic parameters (i.e. effective population size and migration rates).

Besides the wide range of possible uses, microsatellite markers can be transferred across closely related species (Beja-Pereira *et al.* 2004) what makes them economically appealing.

Genomic approaches in conservation are becoming more common, with the possibility of using thousands of SNPs (single nucleotide polymorphism) in order to estimate more accurately a variety of important parameters in conservation (Allendorf 2010). However, the genomics age is on hold on what concerns its use on noninvasive samples. The low quality and quantity of the DNA extract still precludes the direct use of modern genome-wide sequencing or genotyping techniques. Although this limitation can be overcome by using capture arrays or DNA enrichment, the costs of such methodologies are way too expensive to be used at the population level studies.

1.6 Objectives and Structure of the Thesis

The main objectives of this thesis were:

1. **To reassess the evolutionary history of the *Equus* genus**, using DNA sequence analyses to reconstruct phylogenetic relationships among African and Asiatic wild asses (species and subspecies).
2. **To assess levels of genetic diversity and phylogeographic patterns in African and Asiatic wild ass populations**, across a vast geographical distribution range.
3. **To develop or adapt a set of nuclear molecular markers (microsatellites)**, that would perform well on poor quality DNA (i.e., as is the DNA from non-invasive samples), and provide enough information for individual identification, population assignment and species differentiation.
4. **To genotype a representative number of non-invasive samples of extant African wild ass populations** in order to assess key genetic and demographic parameters for the conservation of the African wild ass such as effective population size, gene-flow,

demographic dynamics, population structuring, and degree of inbreeding with the domestic form.

5. **To clarify the contribution of the African wild ass to donkey ancestry and domestication**.
6. **To assess levels of genetic diversity in domestic donkey populations** from geographical areas of interest, in order to identify putative centers of origin for this species.

This thesis is organized in six chapters, including the General Introduction (present chapter).
In Chapter II we addressed Objectives 1 and 2. Within this chapter we reassessed the phylogeny of the *Equus* genus, by looking at relationships among wild-ass species. Samples were collected from ten wild ass populations, comprising the three taxonomically distinct species: *Equus africanus*, *Equus hemionus* and *Equus kiang*. Samples were collected across a vast area of these species' distribution ranges, using a non-invasive methodology. Two mitochondrial genes were used in order to assess levels of genetic diversity, conduct phylogeographic analyses and clarify the evolutionary history of wild ass species.
Obtained results were organized in one article.

Article 1: Rosenbom, S, Costa V., Chen, S., Khalatbari, L., Yusefi, H., Abdukadir, A., Yangzom, C., Moehlman, P., Beja-Pereira, A. *Reassessing the evolutionary history of ass-like equids: insights from variation in contemporary extant populations*. Molecular Phylogenetics and Evolution 85 (2015), 88–96

Objectives 3 and 4 were addressed in Chapter III. The use of non-invasive sampling in population genetics studies of endangered species is the only ethically recommended approach. Given the many technical challenges in the use of these methodologies, a pilot study is recommended in order to assess levels of polymorphism and Mendelian transmission of chosen microsatellite markers and to detect and quantify genotyping errors. We conducted this pilot study by choosing and amplifying a set of 15 microsatellites, isolated from the horse genome, for 22 captive African wild ass individuals. Subsequently, we were able to apply a sub set of 10 microsatellite markers, to a significant number of non-invasive samples from extant African wild ass populations in Ethiopia and Eritrea. MtDNA sequences were also obtained in order to assess the maternal origin of sampled individuals. Obtained results allowed us to assess levels of genetic diversity,

population structure, and migration among extant populations and to detect potential hybridization between the wild and the domestic forms. Results were organized in two articles:

Article 2 – Rosenbom, S., Costa, V., Steck, B. Moehlman, P., Beja-Pereira, A. *Cross-species genetic markers: a useful tool to study the world's most threatened wild equid—Equus africanus*. European Journal of Wildlife Research, 58, 609-613

Article 3 – Rosenbom, S., Kebede, F., Teclai, R., Yohannes, H. Hagos, F. Moehlman, P. and Beja-Pereira, A. *Non-invasive genetic assessment of critically endangered African wild ass (Equus africanus) populations*. In preparation

In Chapter IV we assessed Objectives 5 and 6. The African wild ass is a unique species, from an evolutionary perspective, because it is the living ancestor of a domesticated animal (the domestic donkey). In this way, it is possible to assess the contribution of the African wild ass, to the domestic donkey genetic pool. We have analyzed mitochondrial variation of extant African wild ass populations as well as ancient and historic samples of the domestic donkey, from geographical regions of interest, in order to assess the role of different African wild ass lineages in the domestication process. Additionally, we have conducted a genetic diversity assessment, using a set of microsatellite loci, in domestic donkey populations from geographic regions of interest, in order to identify putative centers of origin of the domestic donkey.
Obtained results were organized in two articles:

Article 4 – Kimura, B., Marshall, F., Chen, S., Rosenbom, S., Moehlman, P., Tuross, N., Sabin, R., Peters, J., Barich, B., Yohannes, H., Kebede, F., Teclai, R., Beja-Pereira, A. amd Mulligan, C. *Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication*. Proceedings of the Royal Society B, 278 (1702), 50-57

Article 5 – Rosenbom, S., Costa, V.Al-Araimi, N., Rukavina, D., Abdel-Moneim, A. and Beja-Pereira, A. *Genetic diversity from donkey populations from the putative centers of domestication*. Animal Genetics, 46 (1), 30-36

Chapter V, results from the collaborative work developed in wild equid non-invasive genetics. Grevy's zebra is a wild equid that inhabits in habitats similar to the African wild ass, occupying geographical regions in Ethiopia and Kenya. It also shares a similar level of endangerment with the African wild ass, facing similar problems such as genetic diversity erosion and population fragmentation. We have used a non-invasive sampling methodology and mitochondrial DNA sequencing in order to assess levels of genetic diversity and phylogeographic patterns in Ethiopian populations of the Grevy´s zebra. This work resulted in the publication of one article:

Article 6 – Kebede, S., Rosenbom, S., Khalatbari, L., Moehlman, P., Beja-Pereira, A. and Bekele, A. *Genetic diversity of the Ethiopian Grevy's zebra populations that includes a unique population of the Alledeghi Plain.* Mitochondrial DNA, 2014 (doi: 10.3109/19401736.2014.898276).

Finally, Chapter VI encloses the General Discussion where the main results of this thesis are discussed and the major conclusions are presented. This chapter also contains the Future Perspectives for new investigation pathways set by the results of this thesis, as well as the need to use new sequencing methodologies.

1.7 References

Allendorf F.W., Hohenlohe, P. A., Luikart, G. (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics* **11**, 697-709.

Allendorf F.W., Luikart, G. (2007) *Conservation and the Genetics of Populations* Blackwell Publishing Massachusetts, USA.

Azzaroli A., Napoleone, G. (1982) Magnetostatigraphic investigation of the Upper Sivaliks near Pinjor, India. *Riv. Italiana Paleontologica Stratigrafia*, 739-62.

Beja-Pereira A., Oliveira R., Alves P.C., Schwartz M.K. & Luikart G. (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. *Mol Ecol Resour* **9**, 1279-301.

Beja-Pereira A., Zeyl E.V.E., Ouragh L., Nagash H., Ferrand N., Taberlet P. & Luikart G. (2004) Twenty polymorphic microsatellites in two of North Africa's most threatened ungulates: Gazella dorcas and Ammotragus lervia (Bovidae; Artiodactyla). *Molecular Ecology Notes* **4**, 452-5.

Bennett D. (1980) Stripes do not a Zebra Make, Part I: A Cladistic Analysis of Equus. *Systematic Biology* **29**, 272-87.

Bennett D. (1992) *The Evolution of the Horse: History and Techiques of Study* Elsevier Publishing Co. , New York, USA.

Bonin A., Bellemain E., Bronken Eidesen P., Pompanon F., Brochmann C. & Taberlet P. (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* **13**, 3261-73.

Churcher C.S. (1982) Oldest ass recovered from Olduvai Gorge, Tanzania, and the origin of asses *Journal of Paleontology* **56**, 1124-32.

Churcher C.S., & Richardson, M.L. (1978) Equidae. In: *Evolution of African mammals* (ed. by Maglio VJ, & Cooke, H.R.S.). Harvard University Press, Cambridge

Clark B. & Duncan P. (1992) Asian wild asses - hemiones and kiangs (E. hemionus Pallas and E. kiang Moorcroft). In: *Zebras, Asses, and Horses. An Action Plan for the Conservation of Wild Equids* (ed. by Duncan P). IUCN, Gland, Switzerland.

Clutton-Brock J. (1992) *Horse power: a history of the horse and the donkey in human societies*. Harvard University Press.

Dalquest W.W. (1978) Phylogeny of Amercan horses from Blancan and Pleistocene Age *Annales Zoologici Fennici* **15**, 191-9.

Eisenmann V. (1976) *Equidae from the Shungura Formation* University of Chicago Press Chicago & London.

Eisenmann V. (1985) Les Equidés des gisements de la vallée de l'Omo en Ethiopie (collections françaises) In: *Les Faunes plio-pléistocènes de la basse vallée de l'Omo (Ethiopie)* (pp. 13-55. CNRS, Paris.

Feh C., Shah, N., Rowen, M., Reading, R. and Goyal, S. P. (2002) Status and action plan for the Asiatic wild ass (*Equus hemionus*). In: *Equids: Zebras, Asses and Horses. Status Survey and Conservation Action Plan* (ed. by Moehlman PD). IUCN, Gland, Switzerland.

Forstén A. (1988) Middle Pleistocene replacment of stenoid horses by calloid horses - ecological implications *Palaeogeography, Palaeoclimatology, Palaeoecology* **65**, 23-33.

Forstén A. (1992) Mitochondrial DNA time-table and the evolution of *Equus*: comparison of molecular and paleontological evidence *Annales Zoologici Fennici* **28**, 301-9.

Frankel O.H., Soulé, M.E. (1981) *Conservation and Evolution*. Cambridge University Press, Cambridge, UK.

Frankham R. (2003) Genetics and conservation biology. *Comptes Rendus Biologies* **326, Supplement 1**, 22-9.

Frankham R. (2005) Genetics and extinction. *Biological Conservation* **126**, 131-40.

Frankham R., Ballou, J.D. and Briscoe, D.A. (2002) *Introduction to Conservation Genetics*. University Press, Cambridge, UK.

George M. & Ryder O.A. (1986) Mitochondrial DNA evolution in the genus Equus. *Molecular Biology and Evolution* **3**, 535-46.

Groves C.P.a.M.V. (1967) On some taxonomic problems of Asiatic wild asses; with the description of a new subspecies (Perissodactyla; Equidae). *Z. Saugetierkunde*, 321-55.

Grubb P. (2005) *Order Perissodactyla*. The Johns Hopkins University Press, Baltimore, Maryland, USA.

Hooijer D.A. (1976) *Evolution of the Perissodactyla of the Omo Group deposits* University of Chicago Press, Chicago & London.

Kaminski M. (1979) The biochemical evolution of the horse. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* **63**, 175-8.

Kebede F. (1999) Ecology and conservation of the African wild ass (Equus africanus) in the Danakil, Ethiopia. University of Kent.

Kurtén B.A., E. (1980) *Pleistocene mammals of North America*. Columbia University Press, New York, USA.

Lindsay E.H., Opdyke, N.D. and Johnson, N.M. (1984) Blancan-Hemphillian land mammal ages and late Cenozoic dispersal events. *Annual Review of Earth and Planetary Sciences* **12**, 445-88.

Lindsay E.H.O., N.D and Johnson, N.M. (1980) Pliocene dispersal of the horse Equus and late Cenozoic mammalian dispersal events. *Nature*, 135-8.

MacFadden B.J. (1992) *Fossil Horses - Systematics, Paleobiology and Evolution of the Family Equidae* Cambridge University Press, Cambridge, United Kingdom

MacFadden B.J. (2005) EVOLUTION: Fossil Horses--Evidence for Evolution. *Science* **307**, 1728-30.

Moehlman P.D. (1998) Feral asses (Equus africanus): intraspecific variation in social organization in arid and mesic habitats. *Applied Animal Behaviour Science* **60**, 171-95.

Moehlman P.D. (2002) *Equids: Zebras, Asses and Horses: Status Survey and Conservation Action Plan*

IUCN/SSC Equid Specialist Group Gland, Switzerland and Cambridge, United Kingdom.

Moehlman P.D., Shah N. & Feh C. (2008a) *Equus hemionus*. URL http://www.iucnredlist.org.

Moehlman P.D., Yohannes H., Teclai R. & Kebede F. (2008b) *Equus africanus*. URL http://www.iucnredlist.org.

Oakenfull E.A. & Clegg J.B. (1998) Phylogenetic Relationships Within the Genus Equus and the Evolution of α and θ Globin Genes. *Journal of Molecular Evolution* **47**, 772-83.

Oakenfull E.A., Lim H. & Ryder O. (2000) A survey of equid mitochondrial DNA: Implications for the evolution, genetic diversity and conservation of Equus. *Conservation Genetics* **1**, 341-55.

Orlando L., Ginolhac A., Zhang G., Froese D., Albrechtsen A., Stiller M., Schubert M., Cappellini E., Petersen B., Moltke I., Johnson P.L.F., Fumagalli M., Vilstrup J.T., Raghavan M., Korneliussen T., Malaspinas A.-S., Vogt J., Szklarczyk D., Kelstrup C.D., Vinther J., Dolocan A., Stenderup J., Velazquez A.M.V., Cahill J., Rasmussen M., Wang X., Min J., Zazula G.D., Seguin-Orlando A., Mortensen C., Magnussen K., Thompson J.F., Weinstock J., Gregersen K., Roed K.H., Eisenmann V., Rubin C.J., Miller D.C., Antczak D.F., Bertelsen M.F., Brunak S., Al-Rasheid K.A.S., Ryder O., Andersson L., Mundy J., Krogh A., Gilbert M.T.P., Kjaer K., Sicheritz-Ponten T., Jensen L.J., Olsen J.V., Hofreiter M., Nielsen R., Shapiro B., Wang J. & Willerslev E. (2013) Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**.

Ryder O.A., Epel N.C. & Benirschke K. (1978) Chromosome banding studies of the Equidae. *Cytogenetic and Genome Research* **20**, 323-50.

Ryder O.A., Sparkes R.S., Sparkes M.C. & Clegg J.B. (1979) Hemoglobin polymorphism in Equus przewalskii and E. caballus analyzed by isoelectric focusing. *Comp Biochem Physiol B* **62**, 305-8.

Schaller G.B. (1998) *Wildlife of the Tibetan Steppe*. University of Chicago Press, Chicago, USA.

Shaffer M.L. (1981) Minimum Population Sizes for Species Conservation. *BioScience* **31**, 131-4.

Shah N. (2002) Status and Action Plan for the Kiang (Equus kiang). In: *Status Survey and Conservation Action Plan - Equids: Zebras, Asses and Horses* (ed. by Moehlman PD), Gland, switzerland.

Shah N., St. Louis, A., Huibin, Z., Bleisch, W., van Gruissen, J. & Qureshi, Q. (2008) *Equus kiang*. URL http://www.iucnredlist.org.

Soulé M.E. & Wilcox B.A. (1980) *Conservation Biology - An Evolutionary-Ecological perspective* Sinauer Associates, Massachusetts, USA

Steiner C.C., Mitelberg A., Tursi R. & Ryder O.A. (2012) Molecular phylogeny of extant equids and effects of ancestral polymorphism in resolving species-level phylogenies. *Molecular Phylogenetics and Evolution* **65**, 573-81.

Steiner C.C. & Ryder O.A. (2011) Molecular phylogeny and evolution of the Perissodactyla. *Zoological Journal of the Linnean Society* **163**, 1289-303.

Taberlet P., Griffin S., Goossens B., Questiau S., Manceau V., Escaravage N., Waits L.P. & Bouvet J. (1996) Reliable Genotyping of Samples with Very Low DNA Quantities Using PCR. *Nucleic Acids Res* **24**, 3189-94.

Taberlet P. & Luikart G. (1999) Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society* **68**, 41-55.

Teclai R. (2006) Conservation of the African wild Ass (Equus africanus) on Messir Plateau (Asa-ila), Eritrea: The role of forage availability and diurnal activity pattern during the wet season and beginning of the dry season. University of Kent.

Van Bemmel A.C.V. (1972) Some remarks on the African wild ass. *Zoologische Mededelingen* **47**.

Wan Q.-H., Wu H., Fujihara T. & Fang S.-G. (2004) Which genetic marker for which conservation genetics issue? *ELECTROPHORESIS* **25**, 2165-76.

Winans M.C. (1989) A quantitave study of the North American fossil species of the genus *Equus*. In: *The Evolution of Perissodactyls* (ed. by Prothero DRS, R.M.). Clarendon Press, Oxford.

Chapter II

Evolutionary history of ass-like equids

Article 1

Molecular Phylogenetics and Evolution, 85 (2015), 88-96

Reassessing the evolutionary history of ass-like equids: insights from patterns of genetic variation in contemporary extant populations

Sónia Rosenbom, Vânia Costa, Shanyuan Chen, Leili Khalatbari, Gholam Hosein Yusefi, Ablimit Abdukadir, Chamba Yangzom, Fanuel Kebede, Redae Teclai, Hagos Yohannes, Futsum Hagos, Patricia Moehlman, and Albano Beja-Pereira

Abstract

All extant equid species are grouped in a single genus – *Equus*. Among those, ass-like equids have remained particularly unstudied and their phylogenetic relations were poorly understood, most probably because they inhabit extreme environments in remote geographic areas. To gain further insights into the evolutionary history of ass-like equids, we have used a non-invasive sampling approach to collect representative fecal samples of extant African and Asiatic ass-like equid populations across their distribution range and mitochondrial DNA (mtDNA) sequencing analyses to examine intraspecific genetic diversity and population structure, and to reconstruct phylogenetic relations among wild ass species/subspecies.

Sequence analyses of 410 base pairs of the fast evolving mtDNA control region identified the Asiatic wild ass population of Kalamaili (China) as the one displaying the highest diversity among all wild ass populations. Phylogenetic analyses of complete cytochrome *b* sequences revealed that African and Asiatic wild asses shared a common ancestor approximately 2.3 Mya and that diversification in both groups occurred much latter, probably driven by climatic events during the Pleistocene. Inferred genetic relationships among Asiatic wild ass species do not support *E. kiang* monophyly, highlighting the need of more extensive studies in order to clarify the taxonomic status of species/subspecies belonging to this branch of the *Equus* phylogeny.

These results highlight the importance of re-assessing the evolutionary history of ass-like equid species, and urges to extend studies at the population level to efficiently design conservation and management actions for these threatened species.

Introduction

Equids (i.e. Equidae family) including the extant horses, donkeys or asses, and zebras, and many other extinct horse-like species have been the subject of numerous studies over the past two centuries. The rich fossil record from the extinct horse-like species has vividly demonstrated a classic example of long-term evolutionary changes and has also been a theme of debate among academics since the end of the nineteenth century (MacFadden 2005). According to currently accepted IUCN taxonomy (Moehlman 2002), modern equids are represented by the eight extant species of the *Equus* genus: domestic horse (*E. caballus*), Przewalski´s horse (*E. przewalskii*), kiang (*E. kiang*), Asiatic wild ass (*E. hemionus*), African wild ass (*E. africanus*), mountain zebra (*E. zebra*), plains zebra (*E. quagga*) and Grevy´s zebra (*E. grevyi*). Under this taxonomy, the domestic donkey (*E. africanus asinus*) is considered a subspecies of African wild ass. Molecular studies in the past three decades (George & Ryder 1986; Oakenfull & Clegg 1998; Oakenfull *et al.* 2000; Kruger *et al.* 2005; Orlando *et al.* 2006; Orlando *et al.* 2009; Steiner & Ryder 2011; Steiner *et al.* 2012; Vilstrup *et al.* 2013) have focused on resolving the complex phylogeny of all extant equids and understanding their evolutionary history, however this has proved to be a challenging task. Besides the confounding effect of incomplete lineage sorting – as a result of rapid and recent divergence – and the probable past and/or present introgression among sympatric species, available phylogenetic studies on the *Equus* genus have often relied on a limited number of samples mostly obtained from zoo collections or captive individuals with few available information about their exact geographical origin (Oakenfull *et al.* 2000; Steiner & Ryder 2011; Steiner *et al.* 2012)

The ass-like branch of the equid phylogeny including Asiatic (*E. hemionus* and *E. kiang*) and African (*E. africanus*) wild asses remains a particularly understudied group of species.

Historically, it is accepted that the Asiatic wild ass ranged from the Arabian peninsula to the Manchuria in a continuous distribution (Grubb 2005). However, over the time, climatic events and increasing anthropogenic impact on their habitats have fragmented their distribution. As a result these species are now scattered in small and isolated populations that are located in arid or high altitude areas in Iran, Turkmenistan, Mongolia, China, and India.

Limited scope studies based on morphology, coat color, geographic location and chromosomal number have been used to justify the split of the Asiatic wild ass into two distinct species – *E. hemionus* and *E. kiang* (Groves 1974; Bennett 1980; Ryder & Chemnick 1990). Nonetheless, molecular data has provided no support for this distinction, with both mitochondrial and genomic data showing *E. kiang* individuals grouping together in a monophyletic clade inside the wider *E. hemionus* variation (Oakenfull *et al.* 2000; McCue *et al.* 2012; Vilstrup *et al.* 2013). Additionally, according to geographical range, three subspecies of *E. kiang* have been proposed (Groves & Mazák 1967): *E. k. kiang*, *E. k. holdereri*, and *E. k. polyodon*, corresponding to Western kiang, Eastern kiang, and Southern kiang, respectively. The validity for these subspecies designations has also been questioned (Schaller 1998; Shah 2002).

It is of particular interest the question regarding the relative position of the African wild ass in the *Equus* phylogeny. Phylogenetic trees based on mtDNA and nuclear loci (Steiner *et al.* 2012) have either placed this species among zebras (mtDNA) or as the earliest diverging taxon of a monophyletic group that comprises all ass-like equids (nuclear loci). African wild asses are phenotypically variable, with two recognized extant subspecies – the Nubian wild ass (*E. a. africanus*) and the Somali wild ass (*E. a. somaliensis*) (Marshall 2007; Moehlman *et al.* 2008b) – occupying distinct geographic areas. Additional genetic data from extant African wild ass populations is required to further assess this taxonomic designation, however political instability in the remote territories occupied by putative extant Nubian wild ass populations has made this task unachievable.

At a time when four out of the seven extant wild equid species are recorded as threatened in the IUCN Red List and conservation resources are limited, it is critical to have a clear understanding of genetic background of equid species and populations in order to appropriately prioritize conservation actions. The use of conservation units has helped in overcoming the taxonomic riddle that many times precludes the establishment of conservation programs at the species/subspecies levels (Crandall *et al.* 2000). The concept of evolutionary significant unit (ESU) has been debated over time, with different authors emphasizing the importance of adaptive distinctiveness (Crandall *et al.* 2000) over the concept of geographic discrete populations or the existence of reciprocal monophyly among proposed conservation units (Moritz 1994). Modern approaches to the definition of ESUs have reinforced the need to combine data from neutral and adaptive markers in order to achieve optimal management decisions (Palsbøll *et al.* 2007; Funk *et al.* 2012).

Despite the advances in non-invasive methodologies (Beja-Pereira *et al.* 2009) and the growing impact of genomic approaches in conservation genetics (Allendorf *et al.* 2010), quantifying divergence at candidate adaptive loci is still a difficult task when using low quality/quantity DNA samples. In such cases, relying on phylogeographic data to define units of conservation might be a necessary first step in an ongoing and complex process.

To further clarify questions regarding the evolutionary history, taxonomy and conservation of the ass-like equid group we conducted a comprehensive study integrating phylogeographic and phylogenetic methodologies, using noninvasive sampling to obtain mtDNA sequences from extant natural populations of African and Asiatic wild asses. We have analyzed samples from 10 populations (see Table 1) representing the African and Asiatic taxa across its entire distribution range (Fig. 1A).

Given the overall decrease in wild equid populations (IUCN 2014) and their level of endangerment it is of crucial importance to have an accurate knowledge about current levels of diversity among extant populations, as well as clarifying the taxonomic status of species. Such actions, along with more reliable information about the ecology of populations, will serve as an important step for prioritizing scientifically based conservation actions, such as the definition of ESUs (Moritz 1994). This is critical for the prevention of further losses in the evolutionary potential of wild equid species.

Materials and methods

Sample collection

We collected fecal samples from three African and seven Asiatic wild ass populations, representing three species: two from Asia (*E. hemionus* and *E. kiang*), and one from Africa (*E. africanus*) (Table 1, Fig. 1). Fecal samples were collected in the field and placed in individual bags. The geographical location (GPS) of each sample was recorded as well as any other relevant information. Samples were dried naturally and stored at room temperature until further processing.

DNA extraction and PCR amplification

To minimize potential contamination issues inherent to non-invasive samples, DNA extraction was carried on a laminar flux chamber, physically separated from the PCR room. The samples

were processed in batches with a maximum of 16 samples per set. All material used during the extraction process was sterilized between sample processing. In each batch of sample DNA extraction, a negative control containing all reagents but not the sample was included to detect contaminations.

DNA extraction was carried using an adapted protocol from JetQuick™ Tissue DNA Spin Kit (Genomed, GmbH). Briefly, the modifications to the standard procedure consist of a pre-treatment with extraction buffer and proteinase K, followed by the use of an inhibitEX® tablet (QIAGEN, GmbH, Hilden). Primers Donk_F (CCC AAG GAC TAT CAA GGA AG) and CR_2R (GGA ATG GCC CTG AAG AAA G) were used to amplify a 440bp fragment of the hypervariable region 1 (HVRI) of the mtDNA control region. Complete cytochrome *b* (Cyt *b*) sequences (1140bp) were obtained by amplifying two overlapping fragments with the following primers: EaCB32F (AAG AAC ACT AAT GAC AAA CAT CC), EaCB687R (GGT GGA ATG GGA TTT TGT C), EaCB625F (TCC ATC TAC TAT TCC TCC ACG) and EaCB997R (CAA GAC CAG GGT AAT GTG TG). PCRs were performed in a 20 µl volume containing 1× PCR buffer, 2.5 mM MgCl2, 0.2 mM dNTPs, 0.6 µM each primer, 0.3 U of PlatinumTaq DNA Polymerase (Invitrogen), and variable amounts of genomic DNA. The PCR mixture underwent 10 min at 95°C, 40 cycles of 45 sec. at 95°C, 60 s at 55°C, 45 s at 72°C, and a final 20 min at 72°C on a GeneAmp PCR System 9700 (Applied Biosystems). Finally, the amplified products were purified and sequenced in the High-Throughput Genomics Unit, University of Washington.

Data analyses

Genetic diversity and population structure

Sequence trace files were edited in DNAStar 7.1 (DNAStar Inc., Madison, WI) and aligned by software Mega version 5.1 (Tamura *et al.* 2011). Sequences from the HVRI of the mtDNA control region were used to calculate genetic diversity parameters including Watterson's theta, haplotype and nucleotide diversity for each population, using DnaSP v5.10 software (Librado & Rozas 2009). A neutrality test (Tajima´s D) was performed using the same software. Additionally 33 previously published Somali wild ass sequences from Eritrea were downloaded from GenBank (supplementary material – table S1).

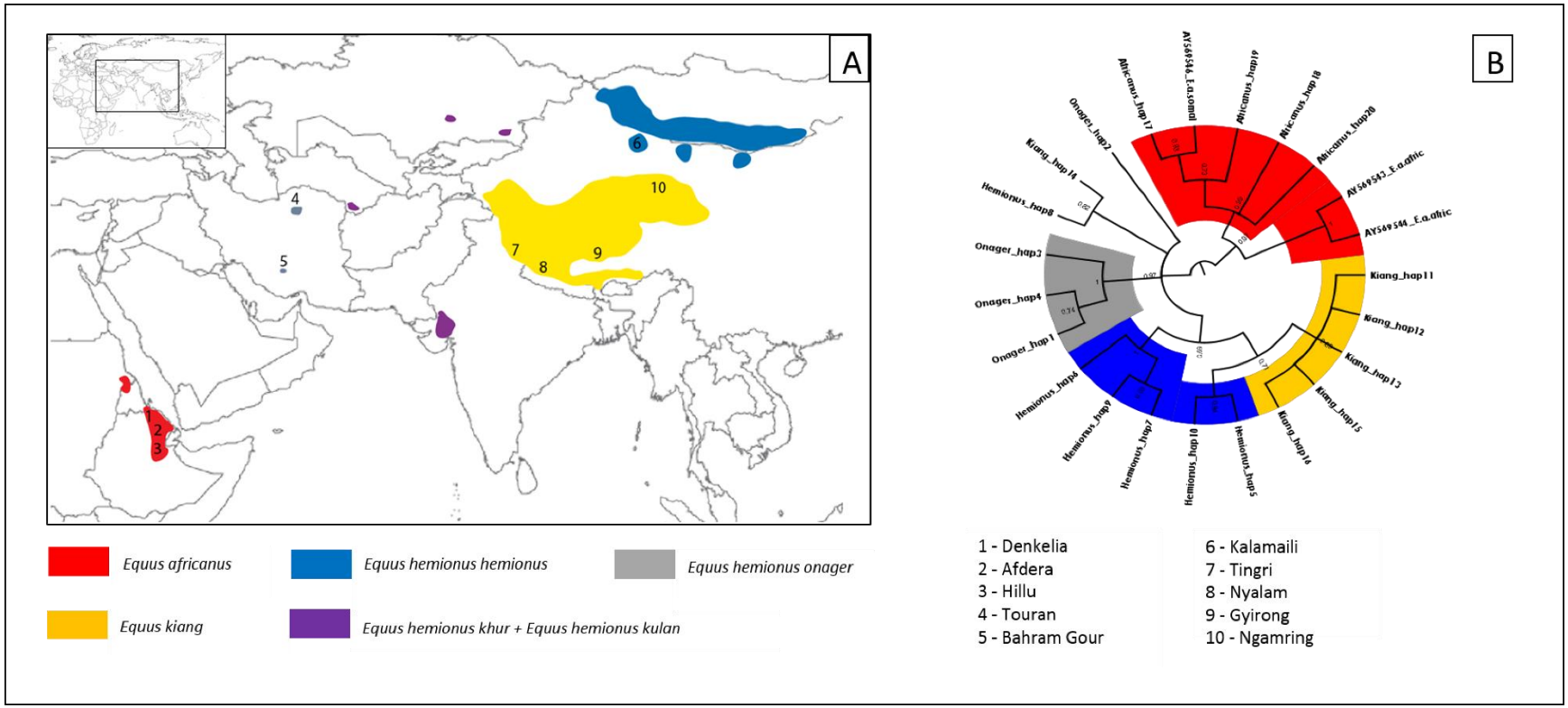


Figure 1 - Current distribution ranges of wild-ass species and location of sampled populations (A). Bayesian tree of obtained control region haplotypes (B). Sequences AY569543, AY569544, AY569546 were retrieved from GenBank).

Geographical structuring of the control region haplotypes was assessed by building a network, for each assumed species, using software NETWORK v4.6 (Bandelt 1999).

To evaluate significant geographic divisions of hypothesized a priori species/ subspecies, we have used hierarchical analyses of molecular variance (AMOVA, Excoffier et al. 1992) in ARLEQUIN v3.5 (Excoffier & Lischer 2010). This analysis divides total variance into variance components via differences among groups (φ_{CT}), among populations within groups (φ_{SC}) and within populations (φ_{ST}). We have tested different population groupings based on both "a priori" taxonomic criteria and obtained results on our phylogenetic analyses. We assumed that the best geographic subdivisions were significantly different from random distributions and expected that the optimal genetic division of species/subspecies will maximize the between-group variance (φ_{CT}) compared to the within-group component (φ_{SC}).

Phylogenetic analyses and molecular dating

To clarify the existence of reciprocal monophyly among wild ass species/subspecies a phylogenetic tree of the control region haplotypes was reconstructed using Bayesian analysis with MrBAYES v3.2.1 (Ronquist *et al.* 2012). Individuals belonging to the most divergent HVRI haplotypes were chosen for complete sequencing of the slower evolving Cyt *b* gene for subsequent phylogenetic analyses and molecular dating of radiation events in equid species. Phylogenetic relationships among newly obtained Cyt *b* haplotypes and previously published equid sequences (supplementary material Table S1) were inferred by using the same software. jModelTest v 2.1.3 (Guindon & Gascuel 2003; Darriba 2012) was used to select the best fitting model of molecular evolution according to the Akaike information criterion (AIC), for both sequence sets. The prior best-fitting nucleotide substitution models for the two data sets were the GTR + G + I and the HKY + G + I, for the HVRI and Cyt *b* sequences, respectively. Bayesian analyses were performed equally for both sequence sets; two independent analyses starting from different random trees were performed, and four MCMC chains were run for 50 million generations with sampling every 1000 generations. Twenty-five percent of the trees were discarded as burn-in, after checking for convergence.

The time to the most recent common ancestors (TMRCAs) of the major clades obtained in the phylogenetic analysis of complete Cyt *b* haplotypes, was estimated using a Bayesian phylogenetic framework implemented in BEAST v1.7.5 (Drummond *et al.* 2012). We have ran three different analyses, assuming a relaxed uncorrelated lognormal molecular clock and a Yule

process of speciation. We have chosen to use only internal calibration points according to the suggestion of a recent study by Vilstrup *et al.* (2013). In the first analyses we have used as a calibration point for the molecular clock the emergence of the *Equus* genus (normal prior distribution centered at 4.0 My; 3-5 My 95% CI) based on the paleontological records for the monodactyle *Plesippus simplicidens*, recognized by some as the earliest fossil of the genus *Equus* (MacFadden & Carranza-Castaneda 2002). This calibration point is further supported by recent phylogenomic studies (Orlando *et al.* 2013; Vilstrup *et al.* 2013) that point towards the origin of all extant equids between 4.0-4.5 Mya.

For the second analyses we have used an alternative calibration point; the emergence of the Plains zebra lineage (normal prior distribution centered at 0.7 My; 0.6-0.8 My 95% CI), according to the fossil record of *E. mauritanicus* (Eisenmann 1979, 1980).

Finally we have ran a third analyses using both previously reported calibration points and described normal prior distributions.

The nucleotide substitution model HKY+ G + I (as in the phylogenetic analysis) was used in MCMC analysis. Parameters were sampled at every 1000 generations over a total of 250 million generations, with 25% generations discarded as burn-in. Convergence of the sampled parameters was checked using TRACER v1.5 (Rambaut & A.J. 2007).

Monophyly was constrained for species represented in the analyses by more than one individual, according to obtained results of the Bayesian phylogenetic analyses.

Demographic dynamics

Bayesian coalescent-based methodology and MCMC sampling procedures implemented in software BEAST v1.7.5 (Drummond *et al.* 2012) were used to estimate the posterior distribution of population size, potentially as far back as TMCRA of a set of samples and given a determined demographic model. We chose the Bayesian skyline model (BSP) which is a piecewise-constant model of population size that allows different demographic scenarios (Drummond *et al.* 2005). This approach was used to make inferences about the demographic history of African and Asiatic wild asses, from mtDNA HVRI sequences, incorporating credibility intervals for the estimated effective population size at every point in time, which accounts for both phylogenetic and coalescent uncertainty. A normal distribution for the strict molecular clock prior was set with 95% of the probability density between 2.6×10^{-8} and 4.6×10^{-8} (mean 3.6×10^{-8} per generation) based on a study using the mtDNA control region of domestic donkeys and African wild asses

(Kimura *et al.* 2011). Other parameters within the BSP model were given uniform distributions, allowing wide range variation. African wild ass samples were divided into Ethiopian and Eritrean populations and ran separately for 10^7 iterations with trees sampled every 1000 iterations. To assess the robustness of the parameter estimates, three independent chains were run with identical settings and combined into a composite chain with 2.7×10^7 states by using software LOGCOMBINER v1.4.7 (Drummond & Rambaut 2007). The same approach was used for Iranian and Chinese wild ass populations and ran the same settings as mentioned above. In both analyses, 10% of the iterations were discarded as burn-in throughout. Log-files were analyzed in TRACER v1.5 (Rambaut & A.J. 2007) and effective sample sizes (ESS) were used to evaluate MCMC convergence within chains.

Results

Genetic diversity and population structure

Analyses of 410bp of the mtDNA HVRI region revealed 55 segregating sites among the 207 samples of African and Asiatic wild ass species, defining a total of 19 haplotypes. Obtained values for both nucleotide (π, Nei 1987) and haplotype (*h*) diversities were higher in the Asiatic wild ass population of Kalimaili (China), that represents the most widespread population of *E. hemionus*, ranging from south-eastern Mongolia to north-western China. Both populations of *E. h. onager* in Iran revealed overall lower values of nucleotide diversity (0.0021 ± 0.0006 and 0.0035 ± 0.0030 for the Touran and Bahram Gour populations, respectively). Tibetan populations of *E. kiang* presented variable values of nucleotide diversity (from 0 to 0.0052 ± 0.0013 for Tingri and Nyalam populations, respectively). African wild ass populations in Ethiopia and Eritrea present vastly different diversity patterns, with nucleotide and haplotype diversity presenting considerably higher values in the Eritrean population ($\pi = 0.0082 \pm 0.0010$ and $h = 0.711 \pm 0.045$), when compared to both Ethiopian populations. Obtained values of Watterson´s theta (θ_W) were overall in line with nucleotide and haplotype diversity, with the exception of the Bahram Gour population, which showed high values for this parameter and moderate low values of haplotype and nucleotide diversities. Obtained results for Tajima's D revealed a negative and highly significant value (Tajima's D = -2.48881; $P < 0.001$), suggesting that the Bahram Gour population might have recently begun to expand.

Geographic structuring of haplotypes revealed marked differences among Asiatic and African wild ass species (Fig. 2-II). Analyses of the *E. hemionus* haplotype network showed no haplotype

sharing between populations in Iran and China, however there was no clear differentiation among haplotypes belonging to both subspecies (Fig. 2-IIB), with one *E. h. onager* haplotype (EH2) clustering among *E. h. hemionus* haplotypes, what might be a consequence of a recent and ongoing process of differentiation between *E. h. hemionus* and *E. h. onager* subspecies. *E. kiang* haplotype network revealed a marked east/west geographical structuring (Fig. 2-IIC), with the most extreme populations of Tingri and Ngamring being represented by the two most divergent haplotypes and the southern population of Nyalam revealing two new unshared haplotypes. Clear differentiation among available *E. a. africanus* and obtained *E. a. somaliensis* haplotypes was found, however three out of four haplotypes observed in the *E. a. somaliensis* were shared between populations in Eritrea and Ethiopia revealing no geographical structure for this subspecies.

Table 1 - Summary statistics of sampled wild ass populations. The table includes country of origin, population and number of samples. Nucleotide diversity (π), haplotype diversity (h), the population parameter θ_W (Watterson, 1975) and Tajima´s D were calculated using mtDNA control region sequences

Species/Subspecies	Country	Population	n	Number of Haplotypes	Number of segregating sites	*h*	π	θ_W	Tajima's D
Equus africanus somaliensis	**Ethiopia**	**Afdera**	**41**	**3**	**8**	**0.261±0.085**	**0.0032±0.0013**	**0.0045±0.0020**	**-0.823**
		Hillu	**6**	**2**	**6**	**0.333±0.215**	**0.0049±0.0032**	**0.0064±0.0038**	**-1.367**
	Eritrea	**Denkelia**	**38**	**4**	**9**	**0.711±0.045**	**0.0082±0.0010**	**0.0052±0.0023**	**1.665**
Equus hemionus onager	**Iran**	**Touran**	**14**	**2**	**2**	**0.440±0.112**	**0.0021±0.0006**	**0.0015±0.0012**	**1.079**
		Bahram Gour	**28**	**3**	**19**	**0.140±0.087**	**0.0035±0.0030**	**0.0119±0.0045**	**-2.488***
Equus hemionus hemionus	**China**	**Kalamaili**	**54**	**6**	**24**	**0.767±0.038**	**0.0173±0.0021**	**0.0129±0.0043**	**1.104**
Equus kiang	**China**	**Nyalam**	**9**	**3**	**5**	**0.667±0.132**	**0.0052±0.0013**	**0.0045±0.0026**	**0.625**
		Tingri	**8**	**1**	**0**	**0.000±0.000**	**0.0000±0.0000**	**0.0000±0.0000**	**nc**
		Gyirong	**5**	**2**	**3**	**0.4±0.237**	**0.0029±0.0017**	**0.0035±0.0025**	**-1.048**
		Ngamring	**4**	**2**	**1**	**0.5±0.265**	**0.0012±0.0007**	**0.0013±0.0013**	**-0.612**

AMOVA results, showed higher among group variation in two out the four tested models (Table 2). Model C reflects currently accepted specific taxonomy, grouping Iranian and Chinese populations of *E. hemionus*; however this model performed poorly when compared to the alternative Model D, grouping Chinese populations belonging to both *E. kiang* and *E. hemionus*

species (Table 2). Model B tested uniquely the geographic criterion, with four clearly defined groups, corresponding to *E. africanus*, *E. h. onager*, *E. h. hemionus* and *E. kiang* populations. This model had the best performance explaining 85.1% of between-group variation and simultaneously presenting the lowest within group variation (7.4%).

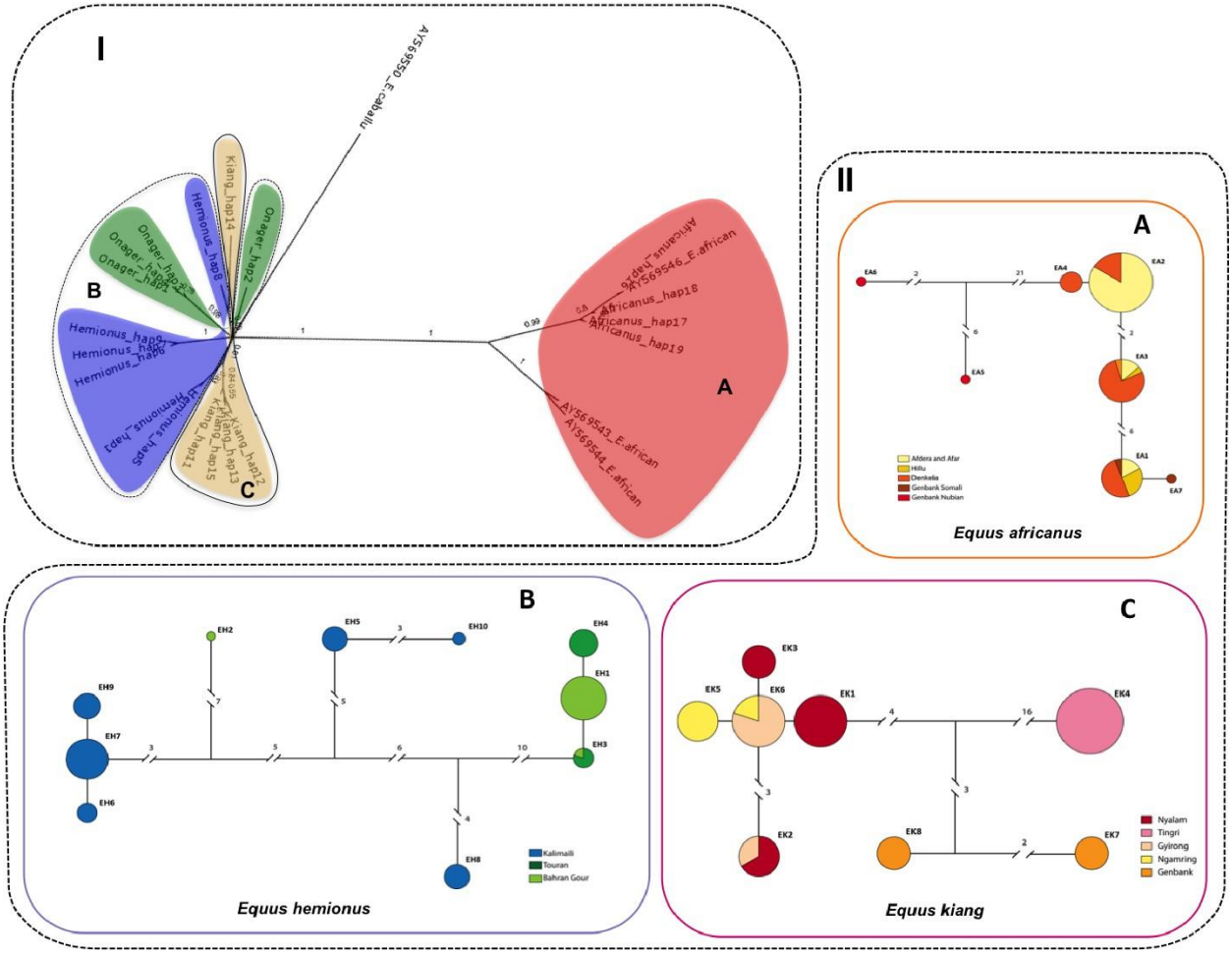


Figure 2 - Phylogenetic relationships between ass-like equid species. I: Rooted phylogenetic network of wild ass species. Numbers above lines are Bayesian posterior probabilities. II: Median-joining mtDNA HVRI haplotype networks for each wild ass species – *Equus africanus* (A), *Equus hemionus* (B) and *Equus kiang* (C).

Table 2 - Analysis of molecular variance for four models of groupings according to accepted taxonomic criteria and obtained phylogenetic results

Model	Hypothesized groups	% Among populations	% Among populations within groups	% Within populations	$P\varphi_{CT}$
A	[Afdera, Hillu, Asmera] [Kalamali, Touran, Bahram Gour, Ngamring, Nyalam, Gyirong, Tingri]	78.13	16.06	5.82	0.01173
B	[Afdera, Hillu, Asmera] [Kalamali] [Touran, Bahram Gour] [Ngamring, Nyalam, Gyirong, Tingri]	85.13	7.36	7.51	0.00196
C	[Afdera, Hillu, Asmera] [Kalamali, Touran, Bahram Gour] [Ngamring, Nyalam, Gyirong, Tingri]	75.95	17.24	6.80	< 0.0001
D	[Afdera, Hillu, Asmera] [Touran, Bahram Gour] [Kalamali, Ngamring, Nyalam, Gyirong, Tingri]	83.10	9.92	6.98	0.0009

Phylogenetic analyses and molecular dating

Bayesian analyses of obtained wild ass HVRI haplotypes revealed two differently supported clades. The clade grouping African wild ass haplotypes, showed high posterior probability values (Fig. 1B), and haplotypes belonging to the two African wild ass lineages clustered in different branches with high support values, validating the hypotheses of two well delimited subspecies – *E. a. africanus* and *E. a. somaliensis* – in opposition to the hypothesis of one genetically uniform species. Asiatic wild ass haplotypes were grouped in a low support clade and reciprocal monophyly among proposed species/subspecies – *E. h. hemionus*, *E. h. onager* and *E. kiang* – was not recovered from the analyses. In fact, although support values for the obtained branches among Asiatic wild asses were overall low and could not fully resolve relationships among these species/subspecies, two major clades can be identified; one grouping *E. h. onager* haplotypes and the other grouping together *E. kiang* and *E. h. hemionus*. These results are not in line with other phylogenetic studies (Oakenfull *et al.* 2000; Steiner *et al.* 2012; Vilstrup *et al.* 2013) in which *hemionus* subspecies clustered together and *E. kiang* haplotypes formed a monophyletic clade inside the wider variation of *E. hemionus*.

Similar results were obtained in Bayesian analyses of Cyt *b* sequences, with two well-supported clades separating African and Asiatic wild ass haplotypes (Fig. 3 and Fig. S1). In the Asiatic wild ass clade it was also possible to retrieve two well supported branches that separated *E. h. onager* from *E. kiang* and *E. h. hemionus* haplotypes. Bayesian phylogenetic analyses of HVRI and Cyt *b* haplotypes are overall concordant and corroborate the hypotheses of *E. h. onager* as the most differentiated population among studied Asiatic wild ass populations.

TMCRA of the major clades obtained by phylogenetic Bayesian analyses of Cyt *b* sequences, using different calibration points within the Equidae phylogeny, were overall concordant (Table 3). Analyses incorporating the *Equus* emergence as a calibration point, revealed a shift in the time range for the arising of the genus, towards the lower limit of the prior with a geometric mean of 3.71 Mya (2.69-4.75; 95% CI). In the analyses incorporating only the calibration point corresponding to the emergence of the Plains zebra lineage, this shift was more accentuated with TMCRA of all extant equids being set at 3.18 Mya (1.4-5.0; 95% CI).

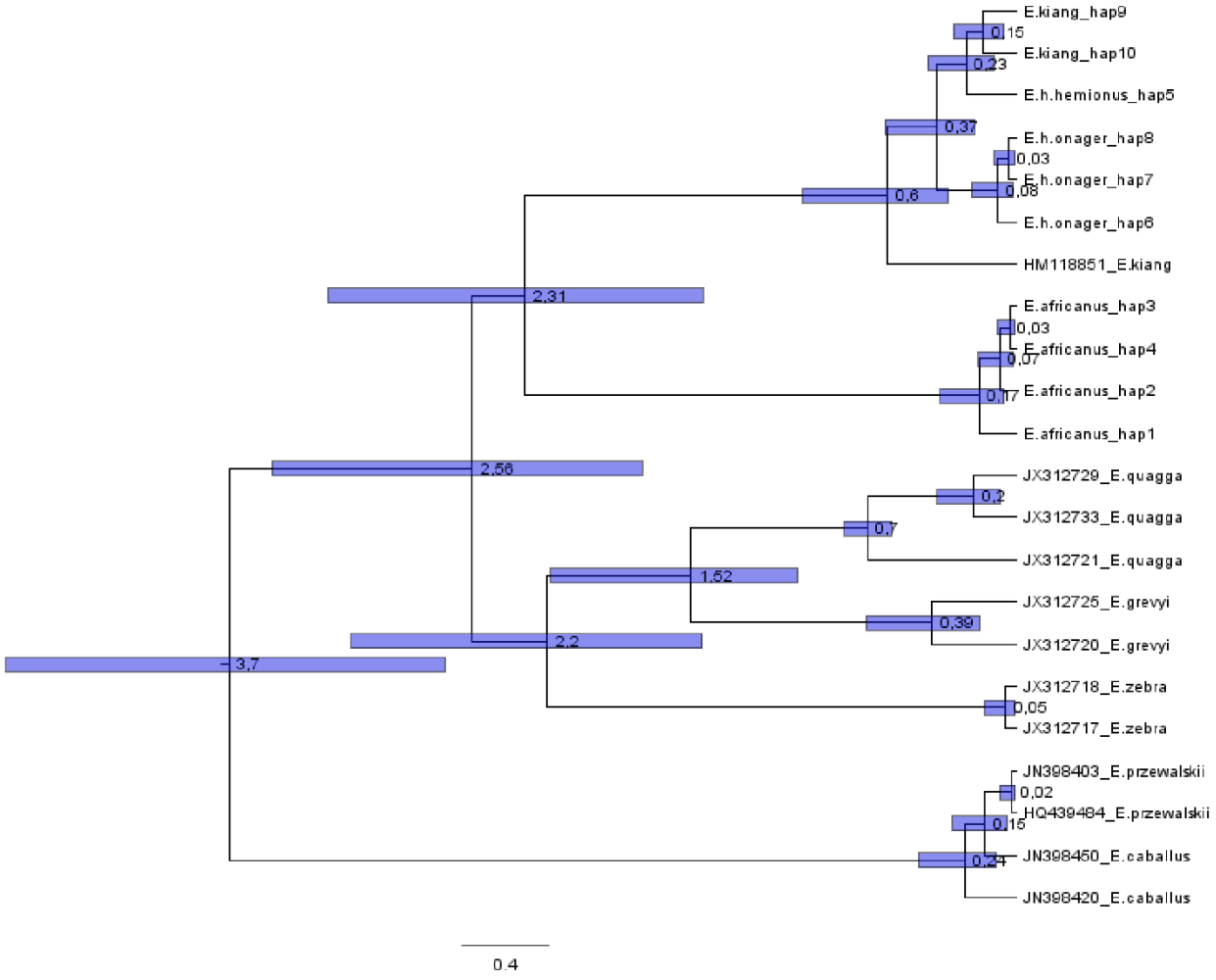


Figure 3 - Node ages, in million years, estimated by BEAST analyses of complete cytochrome *b* sequences, using two internal calibration points from the *Equus* fossil record.

Table 3 - Time for the most common recent ancestor (TMCRA) of main branches obtained by Bayesian phylogenetic inference of cytochrome *b* sequences (Cyt *b*). All dates are in million years with 95% confidence interval given in brackets

	Calibration point 1	Calibration point 2	Calibration point 1+2
TMCRA *Equus* (all species)	3.69 (2.42-4.92; 95% CI)	3.18 (1.4-5.0; 95% CI)	3.71 (2.69-4.75; 95% CI)
TMCRA Zebras (*E.quagga+E.zebra+E.grevyi*)	2.29 (1.21-3.46; 95% CI)	2.02 (1.08-3.08; 95% CI)	2.25 (1.45-3.1; 95% CI)
TMCRA Asses (*E.africanus+E.hemionus+E.kiang*)	2.37 (1.26-3.55; 95% CI)	2.08 (0.99-3.27; 95% CI)	2.34 (1.47-3.24; 95% CI)
TMCRA Horses (*E.caballus+E.prezwalskii*)	0.27 (0.08-0.5; 95% CI)	0.24 (0.07-0.46; 95% CI)	0.26 (0.1-0.46; 95% CI)
TMCRA African wild ass (*E.africanus*)	0.2 (0.04-0.4; 95% CI)	0.18 (0.04-0.36;95% CI)	0.19 (0.05-0.36; 95% CI)
TMCRA Asiatic wild ass (*E.hemionus+E.kiang*)	0.65 (0.27-1.1; 95% CI)	0.58 (0.25-0.98; 95% CI)	0.63 (0.32-1.0; 95% CI)
TMCRA Plain zebras (*E. quagga*)	0.73 (0.32-1.22; 95% CI)	0.68 (0.56-0.8; 95% CI)	0.7 (0.58-0.81; 95% CI)

Obtained mean value for TMCRA of the wild-ass branch of the phylogeny varied from 2.08 Mya (0.99-3.27; 95% CI) to 2.37 Mya (1.26-3.55; 95% CI), depending on the calibration point used for molecular dating analyses (Table 3). Both TMCRA for the African and the Asiatic wild ass branches of the *Equus* phylogeny showed little variation, according to the calibration point used (Table 3); TMCRA of the African wild ass at approximately 200,000 years ago, is considerably younger than that for the Asiatic wild ass branch at about 630,000 years ago. These results support climatic events occurring during the Pleistocene as the major driving force in the differentiation processes in both groups of animals.

Demographic dynamics

Demographic dynamics of Asiatic and African wild ass populations revealed two different scenarios. Focal populations of Asiatic wild ass in Iran and China showed comparable demographic histories, with the parameter value $N_e g$, which stands as a proxy for maternal effective population size, maintaining a relative stability until 25,000 years ago, when a population decline was detected (Fig S2-C and S2-D). In contrast African wild ass population in Ethiopia and Eritrea remained stable over time, revealing lower effective population sizes (Fig S2-A and S2-B), that corroborates the hypotheses of historic low population sizes and the limited distribution of this species, when compared to the closely related Asiatic wild ass.

Discussion

All species belonging to the *Equus* genus have a common and recent origin, with mtDNA studies identifying two deep clades, namely, the caballines and the zebras/asses (MacFadden 2005). These deep clades have split approximately 3 million years ago (Mya) in North America and subsequently dispersed into the Old World (MacFadden 2005), with newly obtained genomic data suggesting that the *Equus* lineage giving rise to all contemporary horses, zebras and donkeys originated 4.0–4.5 Mya (Orlando *et al.* 2013; Vilstrup *et al.* 2013). The rich record of Pliocene equids of North America provided no support for more than one species by the early Blancan of North America, however by late Blancan (2.5-3.0 Mya) the first signs of differentiation start to appear. Scarce paleontological data from faunal assemblies refer to the possibility of a slender-limbed species existing in North America at that period of time, possibly representing the stem group of wild asses (Azzaroli 1982).

In the present study we have used newly obtained Cyt *b* sequences from extant African and Asiatic species, in order to calculate TMCRA of the wild ass branch. Obtained results showed that African and Asiatic wild asses shared a common ancestor approximately 2.3 Mya (1.26-3.55; 95% CI), supporting the hypotheses of wild asses co-existing with early horses in North America, prior to dispersal to the Old World. These results seem to be in good agreement with a recent study that analyzed mitogenomes from all extant equid lineages which places the common ancestor of the wild-ass branch at 2.6 Mya (Vilstrup *et al.* 2013).

Equus is believed to have dispersed to Eurasia, before the end of the Pliocene, arriving to India by 2.5 Mya, Western Europe simultaneously or even slightly earlier and East Africa at about 2 Mya (Azzaroli, 1992; Lindsay, 1980). If this was the case, then the origin of wild asses and the first *Equus* dispersal events to the Old world happened approximately at the same time range, raising the hypothesis that competition for habitat and resources triggered the first long range dispersal events.

Our Bayesian phylogenetic analyses of Cyt *b* sequences resulted on a well-supported unrooted tree (Fig. S1), in which the African wild branch appears as a sister lineage to Asiatic wild asses. Both African and Asiatic wild ass branches were found monophyletic with high posterior probability values. This pattern is consistent with other studies (McCue *et al.* 2012; Steiner *et al.* 2012; Vilstrup *et al.* 2013) that also support African and Asiatic wild asses as sister species. Among Asiatic wild asses, we were unable to retrieve *E. kiang* monophyly, and instead found *E. h. onager* haplotypes to cluster inside the wider *E.hemionus* variation in a well-supported monophyletic clade. These results are in clear disagreement with recent studies (Steiner *et al.*

2012; Vilstrup *et al.* 2013) that show *E. hemionus* subspecies clustering together in a group divergent from *E. kiang*. Such inconsistencies might be a result of incomplete lineage sorting among Asiatic wild asses, resulting in random phylogenetic reconstructions. Despite incongruences in tree topology, TMCRA of the Asiatic wild branch 630 Kya (thousand years ago) is in clear agreement with that obtained by Vilstrup *et al.* (2013) at 672 Kya. TMCRA of the African wild ass at 200 Kya was much younger than previously published results (Oakenfull *et al.* 2000; Vilstrup *et al.* 2013), however we have used sequences belonging only to the Somali wild ass lineage (*E. a. somaliensis*) what could account for this discrepancy.

African and Asiatic wild ass species have undergone different evolutionary processes. The African wild ass seems to be less successful when compared with the closely related Asiatic ass, showing a historically more restricted distribution, in marginal arid habitats in the Horn of Africa. On the other hand, Asiatic wild asses were widely distributed from the Asiatic Far East (southeastern Mongolia) into the Near Eastern Mediterranean shores and as far south as central Arabian Peninsula and the North of the Indian Subcontinent, thus occupying a vast territory.

Asiatic wild ass populations of *E. hemionus* have been stable until approximately 25,000 years ago, when a population decline is observed (Fig. S2-C and S2-D). This decline coincides with the time of the last glacial maximum (LGM), between 19000 and 26000 years ago, when large mammals vanished from many biogeographic regions, finding refuge or shifting distributions to the southern extreme of their range. Although it has been argued that East Asia has never been covered in ice sheets during the last glaciation events (Zhang *et al.* 2008) mainly due to the monsoon effect, the indirect impact of global climatic change most probably affected all parts of the world (Lister & Stuart 2008). After climatic conditions improved, *E. hemionus* populations would have the capacity to recover, however by then the human population had already occupied many of the areas inhabited by these animals and hunting and habitat loss continued to impose a decreasing trend until current days. Similar results were found for closely related wild horses (Lorenzen *et al.* 2011) that showed a decline in genetic diversity after the LGM, reflecting the impact of expanding human populations in Europe and Asia.

Comparatively, the African wild ass apparently went through different demographic dynamics. Perhaps due to their more restricted distribution in a region of the globe less affected by the climatic events, African wild ass populations have managed to keep stable effective population sizes over time.

The current levels of genetic diversity among wild ass populations reflect a very intricate evolutionary process. For instances, the Iranian populations of *E. h. onager* presented low genetic

diversity as a consequence of the accentuate decrease in numbers during the course of the last decades, essentially due to habitat destruction, overgrazing and poaching (Moehlman *et al.* 2008a). Iranian populations are now separated by more than 700 km, subsisting in isolation what is an added concern for their survival. Diversity levels are considerably higher in the Kalimaili population of *E. h. hemionus*, which represents the most widely distributed subspecies of Asiatic wild asses. Despite the apparent genetic health, *E. h. hemionus* populations have been losing suitable habitat across their current range and populations in Mongolia and China are declining and becoming increasingly isolated. Diversity levels in populations of *E. kiang* in Tibet are highly variable (Table 1), however their distribution across the isolated region of the Tibetan plateau, makes them less vulnerable to human mediated actions. Obtained geographical structure of control region haplotypes in *E. kiang* populations, in a comparably smaller area than the wider Iran-China distribution range of the *hemionus spp*. might reflect the unique features of the habitat occupied by these populations, with altitude and harsh mountain slopes working as geographic barriers and promoting differentiation.

E. kiang and *E. hemionus* have allopatric distribution ranges and share morphological, ecological and behavioral similarities, however on the bases of coat colour and karyotypic dissimilarities, they have been considered different species. Our phylogenetic analyses of Cyt *b* haplotypes revealed a clear pattern of differentiation between *E. h. onager* and the cluster that incorporates *E. kiang* and *E. h. hemionus* haplotypes (Fig. 3 and Fig. S1). AMOVA results also supported these results, with the geographical criterion overpowering the taxonomic groupings of different Asiatic wild ass species/subspecies. The larger percentage of variation justified by differences between groups is obtained when *E. kiang, E. h. hemionus* and *E. h. onager* are placed in different groups (Model B – Table 2) and when *E. kiang* and geographically closest population of *E. h. hemionus* are grouped together (Model D – Table 2).

E. africanus is currently the most threatened equid species, with less than 500 individuals subsisting in Ethiopia and Eritrea (Moehlman *et al.* 2008a). Nubian wild asses (*E. a. africanus*) are extremely rare and may already be extinct, however exploratory trips in the northern regions of Eritrea need to be done. In fact, the lack of extensive sampling in contemporary populations has been an obstacle to fully understand the level of differentiation among African wild asses. By sampling Eritrean and Ethiopian populations of *E. a. somaliensis*, we were able to clearly demonstrate the absence of geographical structure, with populations across the border sharing three of the four obtained HVRI haplotypes. Gene flow among *E. a. somaliensis* populations is possible given the lack of geographical barriers. The lack of geographic structure in HVRI

haplotypes indicates that these populations were most probably connected and gene flow may still occur. Comparison of obtained haplotypes with previously published *E. a. africanus* haplotypes, revealed clear differentiation, reinforcing the hypotheses of two well defined subspecies, corresponding to the Nubian and Somali lineages.

Conclusions

Here, we have shown that by using an integrative approach that combines non-invasive sampling together with phylogenetic and phylogeographic methodologies it is possible to obtain new and more complete insights into the evolutionary history of wild ass species. Besides the important contribution to knowledge on historical processes that led to current genetic variation, this work could serve as a framework for the development of conservation actions in wild ass populations and furthermore call for a reevaluation of taxonomic status in Asiatic wild ass species.

Acknowledgments

This work was funded by FEDER funds though a project co-funded by the EU program COMPETE and the national funds from the Portuguese Foundation for Science and Technology (FCT) (PTDC/BIA-BIC/118107/2010 and FCOMP-01-0124-FEDER-019757) and from another FCT project (PTDC/BIA-BDE/64111/2006). AB-P is supported by FCT through an IF-FCT contract. PDM was supported by EcoHealth Alliance, Basel Zoo, Saint Louis Zoo and Liberec Zoo.

References

Allendorf F.W., Hohenlohe P.A. & Luikart G. (2010) Genomics and the future of conservation genetics. *Nat Rev Genet* **11**, 697-709.

Azzaroli A., Napoleone, G. (1982) Magnetostatigraphic investigation of the Upper Sivaliks near Pinjor, India. *Riv. Italiana Paleontologica Stratigrafia*, 739-62.

Bandelt H.-J., Forster, P., Rohl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**, 37-48.

Beja-Pereira A., Oliveira R., Alves P.C., Schwartz M.K. & Luikart G. (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. *Mol Ecol Resour* **9**, 1279-301.

Bennett D. (1980) Stripes do not a Zebra Make, Part I: A Cladistic Analysis of Equus. *Systematic Biology* **29**, 272-87.

Crandall K.A., Bininda-Emonds O.R.P., Mace G.M. & Wayne R.K. (2000) Considering evolutionary processes in conservation biology. *Trends Ecol Evol* **15**, 290-5.

Darriba D., Doallo, R., Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**.

Drummond A.J., Rambaut A., Shapiro B. & Pybus O.G. (2005) Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. *Mol Biol Evol* **22**, 1185-92.

Drummond A.J., Suchard M.A., Xie D. & Rambaut A. (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**, 1969-73.

Eisenmann V. (1979) Evolutionary characters and phylogeny of the genus Equus (Mammalia, Perissodactyla). *Comptes Rendus Hebdomadaires des Seances de L'Academie des Sciences D* **288**, 497-500.

Eisenmann V. (1980) Les chevaux (*Equus* sensu latu) fossiles et actuels: cranês et dents jugales supérieures *Cah. Paléont.*

Funk W.C., McKay J.K., Hohenlohe P.A. & Allendorf F.W. (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution* **27**, 489-96.

George M. & Ryder O.A. (1986) Mitochondrial DNA evolution in the genus Equus. *Molecular Biology and Evolution* **3**, 535-46.

Groves C.P. (1974) *Horses, asses and zebras in the Wild*. Ralph Curtis Books, Hollywood, Florida, USA.

Groves C.P. & Mazák V. (1967) On some taxonomic problems of Asiatic wild asses; with the description of a new subspecies (Perissodactyla; Equidae). *Z. Saugetierkunde* **32**, 321-55.

Grubb P. (2005) *Order Perissodactyla*. The Johns Hopkins University Press, Baltimore, Maryland, USA.

Guindon S. & Gascuel O. (2003) A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Syst Biol* **52**, 696-704.

IUCN (2014) The IUCN Red List of Threatened Species. Version 2014.3. URL www.iucnredlist.org.

Kimura B., Marshall F.B., Chen S., Rosenbom S., Moehlman P.D., Tuross N., Sabin R.C., Peters J., Barich B., Yohannes H., Kebede F., Teclai R., Beja-Pereira A. & Mulligan C.J. (2011) Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication. *Proc Biol Sci* **278**, 50-7.

Kruger K., Gaillard C., Stranzinger G. & Rieder S. (2005) Phylogenetic analysis and species allocation of individual equids using microsatellite data. *J Anim Breed Genet* **122 Suppl 1**, 78-86.

Librado P. & Rozas J. (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-2.

Lister A.M. & Stuart A.J. (2008) The impact of climate change on large mammal distribution and extinction: Evidence from the last glacial/interglacial transition. *Comptes Rendus Geoscience* **340**, 615-20.

Lorenzen E.D., Nogues-Bravo D., Orlando L., Weinstock J., Binladen J., Marske K.A., Ugan A., Borregaard M.K., Gilbert M.T., Nielsen R., Ho S.Y., Goebel T., Graf K.E., Byers D.,

Stenderup J.T., Rasmussen M., Campos P.F., Leonard J.A., Koepfli K.P., Froese D., Zazula G., Stafford T.W., Jr., Aaris-Sorensen K., Batra P., Haywood A.M., Singarayer J.S., Valdes P.J., Boeskorov G., Burns J.A., Davydov S.P., Haile J., Jenkins D.L., Kosintsev P., Kuznetsova T., Lai X., Martin L.D., McDonald H.G., Mol D., Meldgaard M., Munch K., Stephan E., Sablin M., Sommer R.S., Sipko T., Scott E., Suchard M.A., Tikhonov A., Willerslev R., Wayne R.K., Cooper A., Hofreiter M., Sher A., Shapiro B., Rahbek C. & Willerslev E. (2011) Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature* **479**, 359-64.

MacFadden B.J. (2005) EVOLUTION: Fossil Horses--Evidence for Evolution. *Science* **307**, 1728-30.

MacFadden B.J. & Carranza-Castaneda O. (2002) Cranium of *Dinohippus mexicanus* (Mammalia Equidae) from the early Pliocene (latest Hemphillian) of central Mexico and the origin of *Equus*. *Bull. Florida Museum Nat History*, 163-85.

Marshall F. (2007) *Rethinking Agriculture: Archeological and Ethnoarcheological perspectives*. Left Coast Press, Inc., Walnut Creek, California, USA.

McCue M.E., Bannasch D.L., Petersen J.L., Gurr J., Bailey E., Binns M.M., Distl O., Guerin G., Hasegawa T., Hill E.W., Leeb T., Lindgren G., Penedo M.C., Roed K.H., Ryder O.A., Swinburne J.E., Tozaki T., Valberg S.J., Vaudin M., Lindblad-Toh K., Wade C.M. & Mickelson J.R. (2012) A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. *PLoS Genet* **8**, e1002451.

Moehlman P.D. (2002) *Equids: Zebras, Asses and Horses. Status Survey and Conservation*. IUCN, Gland, Switzerland.

Moehlman P.D., Shah N. & Feh C. (2008a) *Equus hemionus*. In: *IUCN Red List of Threatened Species. Version 2012.2.*

Moehlman P.D., Yohannes H., Teclai R. & Kebede F. (2008b) *Equus africanus*. URL http://www.iucnredlist.org.

Moritz C. (1994) Defining 'Evolutionarily Significant Units' for conservation. *Trends Ecol Evol* **9**, 373-5.

Oakenfull E.A. & Clegg J.B. (1998) Phylogenetic Relationships Within the Genus Equus and the Evolution of α and θ Globin Genes. *Journal of Molecular Evolution* **47**, 772-83.

Oakenfull E.A., Lim H. & Ryder O. (2000) A survey of equid mitochondrial DNA: Implications for the evolution, genetic diversity and conservation of Equus. *Conservation Genetics* **1**, 341-55.

Orlando L., Ginolhac A., Zhang G., Froese D., Albrechtsen A., Stiller M., Schubert M., Cappellini E., Petersen B., Moltke I., Johnson P.L., Fumagalli M., Vilstrup J.T., Raghavan M., Korneliussen T., Malaspinas A.S., Vogt J., Szklarczyk D., Kelstrup C.D., Vinther J., Dolocan A., Stenderup J., Velazquez A.M., Cahill J., Rasmussen M., Wang X., Min J., Zazula G.D., Seguin-Orlando A., Mortensen C., Magnussen K., Thompson J.F., Weinstock J., Gregersen K., Roed K.H., Eisenmann V., Rubin C.J., Miller D.C., Antczak D.F., Bertelsen M.F., Brunak S., Al-Rasheid K.A., Ryder O., Andersson L., Mundy J., Krogh A., Gilbert M.T., Kjaer K., Sicheritz-Ponten T., Jensen L.J., Olsen J.V., Hofreiter M., Nielsen R., Shapiro B., Wang J. & Willerslev E. (2013) Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**, 74-8.

Orlando L., Mashkour M., Burke A., Douady C.J., Eisenmann V. & Hanni C. (2006) Geographic distribution of an extinct equid (Equus hydruntinus: Mammalia, Equidae) revealed by morphological and genetical analyses of fossils. *Mol Ecol* **15**, 2083-93.

Orlando L., Metcalf J.L., Alberdi M.T., Telles-Antunes M., Bonjean D., Otte M., Martin F., Eisenmann V., Mashkour M., Morello F., Prado J.L., Salas-Gismondi R., Shockey B.J., Wrinn P.J., Vasil'ev S.K., Ovodov N.D., Cherry M.I., Hopwood B., Male D., Austin J.J., Hänni C. & Cooper A. (2009) Revising the recent evolutionary history of equids using ancient DNA. *Proc Natl Acad Sci U S A* **106**, 21754-9.

Palsbøll P.J., Bérubé M. & Allendorf F.W. (2007) Identification of management units using population genetic data. *Trends in Ecology & Evolution* **22**, 11-6.

Rambaut A. & A.J. D. (2007) Tracer version 1.4., http://tree.bio.ed.ac.uk/software/tracer/.

Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A. & Huelsenbeck J.P. (2012) MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst Biol* **61**, 539-42.

Ryder O.A. & Chemnick L.G. (1990) Chromosomal and molecular evolution in Asiatic wild asses. *Genetica* **83**, 67-72.

Schaller G.B. (1998) *Wildlife of the Tibetan Steppe*. University of Chicago Press, Chicago, USA.

Shah N. (2002) Status and Action Plan for the Kiang (Equus kiang). In: *Status Survey and Conservation Action Plan - Equids: Zebras, Asses and Horses* (ed. by Moehlman PD), Gland, switzerland.

Steiner C.C., Mitelberg A., Tursi R. & Ryder O.A. (2012) Molecular phylogeny of extant equids and effects of ancestral polymorphism in resolving species-level phylogenies. *Molecular Phylogenetics and Evolution* **65**, 573-81.

Steiner C.C. & Ryder O.A. (2011) Molecular phylogeny and evolution of the Perissodactyla. *Zoological Journal of the Linnean Society* **163**, 1289-303.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M. & Kumar S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-9.

Vilstrup J.T., Seguin-Orlando A., Stiller M., Ginolhac A., Raghavan M., Nielsen S.C., Weinstock J., Froese D., Vasiliev S.K., Ovodov N.D., Clary J., Helgen K.M., Fleischer R.C., Cooper A., Shapiro B. & Orlando L. (2013) Mitochondrial phylogenomics of modern and ancient equids. *PLoS One* **8**, e55950.

Zhang H., Yan J., Zhang G. & Zhou K. (2008) Phylogeography and Demographic History of Chinese Black-Spotted Frog Populations (Pelophylax nigromaculata): Evidence for Independent Refugia Expansion and Secondary Contact. *BMC Evol Biol* **8**, 21.

Supplementary material

Table S1 - Information for GenBank retrieved sequences

Species	GenBank acc.	Ref. paper	Geographical origin
Equus caballus	JN 398420	Achilii et *al*., 2012	n.a.
Equus caballus	JN 398450	Achilii et *al*., 2012	n.a.
Equus przewalskii	JN 398403	Achilii et *al*., 2012	n.a.
Equus przewalskii	HQ 439484	Lippold et *al*., 2010	n.a.
Equus kiang	HM118851	Luo et *al*., 2011	Tibet
Equus zebra	JX 312717	Vilstrup et *al*., 2013	Namibia
Equus zebra	JX312718	Vilstrup et *al*., 2013	Namibia
Equus grevyi	JX312725	Vilstrup et *al*., 2013	Kenya
Equus grevyi	JX312720	Vilstrup et *al*., 2013	Kenya
Equus quagga	JX312721	Vilstrup et *al*., 2013	Kenya
Equus quagga	JX312733	Vilstrup et *al*., 2013	n.a.
Equus quagga	JX312729	Vilstrup et *al*., 2013	n.a.

n.a. - not applicable

Figure S1 - Bayesian phylogenetic tree of Cytochrome *b* sequences. Posterior probabilities are given for each node.

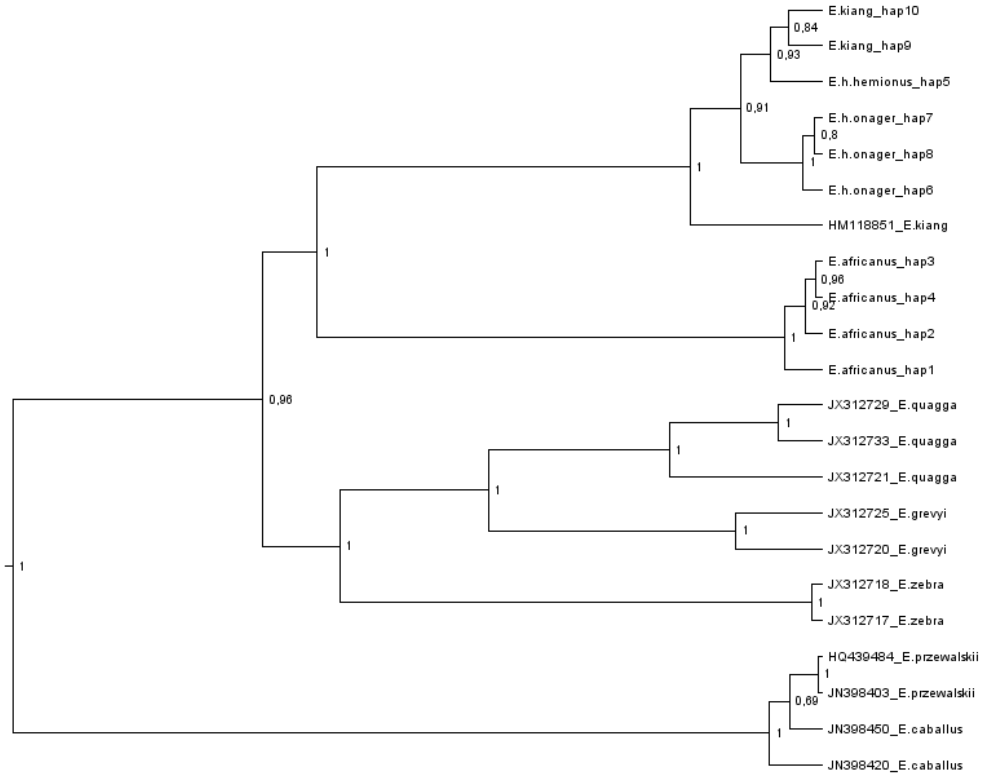
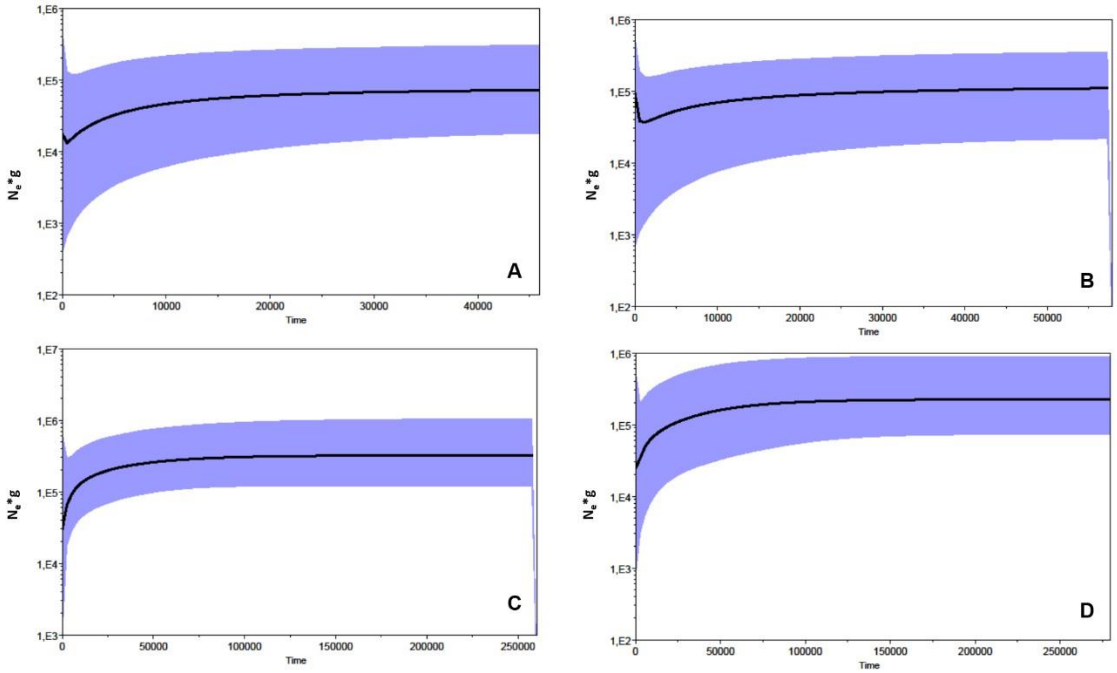


Figure S2 - Bayesian skyline plots of main wild ass populations, calculated from obtained HVRI sequences; A – Ethiopia; B – Eritrea; C – China (Kalimali) and D – Iran (Bahram Gour and Touran).



Chapter III

Conservation genetics of the critically endangered African wild ass (*Equus africanus*): a non-invasive molecular approach

Article 2

European Journal of Wildlife Research, 58, 609-613

Cross-species genetic markers: a useful tool to study the world's most threatened wild equid - Equus africanus

Sónia Rosenbom, Vânia Costa, Beatrice Steck, Patricia Moehlman and Albano Beja-Pereira

Abstract

Once a diverse family, the Equidae family is now reduced to a single genus, *Equus*. From the seven extant species of the genus, the African wild ass (*Equus africanus*) is the most threatened with extinction (last survey indicated 600 individuals). In this work we tested 25 published microsatellite primer pairs isolated from the horse genome on 22 African wild ass (*E. africanus*) individuals from wildlife reserves and zoos. From the 25 loci tested, 15 amplified well and showed moderate allelic richness (5.06, mean number of alleles) and moderately high expected heterozigosity (0.59). Although all possible loci pairs showed no significant gametic disequilibrium (P>0.007), deviations from Hardy–Weinberg proportions were found in 2 out of the 15 analyzed microsatellite loci (AHT5 and VHL20). Here, we propose these polymorphic markers to be used as a standard set in future studies on population and conservation genetics of the African wild ass.

Introduction

During the Pleistocene, equids were the most abundant, medium-sized grazing animals of the grasslands and steppes of Africa, Asia and the Americas. Today there remain only seven species and many of these are at risk.

The African wild ass (*Equus africanus*) is one of the two *Equus* species that is considered critically endangered (Moehlman 2002; Boyd *et al.* 2008; Moehlman 2008). Due to their extremely fragile

conservation status, most of these species will need, in the near future, proper genetic diversity studies in order to establish adequate conservation plans.

The African wild ass stands as a particular important species in the *Equus* genus, not only due to its worrying conservation status but especially because being the wild ancestor of the domestic donkey (Beja-Pereira *et al.* 2004a). It provides a unique link for unravelling marks of the domestication process by comparison of modern and ancestral genomes.

Over the last decades, anthropogenic pressures, together with rapid environmental changes and a possible hybridization with the domestic donkey, have been driving the African wild ass to the brink of extinction. Once distributed across a wider range, north and west into Sudan, Egypt and Libya, the African wild ass is now circumscribed to the Horn of Africa (the desert of Ethiopia, Eritrea and Somalia and possibly Sudan). Although hunting and the commerce of specimens is forbidden for this species, the impact caused by the small effective population size, habitat fragmentation and the possibility of interbreeding of the wild individuals with domestic donkeys is a matter of great concern in terms of genetic health and integrity of the extant wild populations that urges to be investigated.

Microsatellites are the most widely used markers in population genetics because they are highly polymorphic, codominant, abundant throughout the genome and amenable to polymerase chain reaction even in low quality biological samples like the ones resulting from non-invasive sampling techniques (Beja-Pereira *et al.* 2009). Another positive aspect of using microsatellites is the fact that these markers can easily be transferred across closely related taxa and thus be widely and successfully applied in conservation studies. Many studies have already confirmed that microsatellite flanking sequences are often conserved across species and even across families, especially in ungulates (Beja-Pereira *et al.* 2004b). This point is particularly important as it allows overcoming financial and time constrains to clone markers for the threatened equid species.

Material and Methods

DNA extraction and microsatellite genotyping

In this work we tested 25 microsatellites loci that have been previously used in several horse studies - AHT4, AHT5, ASB2, ASB17, ASB23, CA425, COR20, COR58, COR70, COR75, COR90, HMS3, HMS6, HMS7, HMS20, HTG6, HTG7, HTG10, LEX68, LEX74, NVHEQ18, TKY321, UCDEQ457, UM11 and VHL20 - on 22 individuals from wildlife reserves and zoos.

DNA extraction was performed using standard and adapted protocols from Qiagen® DNeasy® Blood and Tissue Kit.

Three blood and five tissue samples were extracted according to the manufacturers' standard protocols. Fourteen stool samples were extracted using the same kit with the following modifications - around 100 mg of the stool surface was removed and washed with 1 mL PBS buffer (137 mM NaCl, 1.7 mM Kcl, 4.3 mM Na2HPO4, 1.47 mMKH2PO4; pH 7.4) and incubated at room temperature under agitation during 30 min. Two hundred microliters of this solution was removed to a new tube, avoiding any solid particles, and the remaining steps were carried according to standard blood protocol for this kit.

The genomic DNA was amplified by polymerase chain reaction (PCR) and the forward primers were modified by end labelling with fluorescent dyes (6-FAM™, VIC®, NED™, PET®) at the 5' end. Each 15-µl reaction consisted of water, DNA, primers and fluorescent labels (0.06 µM primer forward, 0.6 µM primer reverse and dye), dNTPs (30 mM each), 10× buffer [200 mM Tris–HCl (pH 8.4), 500 mM KCl], BSA (0.4 µg/µl), MgCl2 (variable between 1.5 and 3 mM, according to the loci) and Platinum® Taq DNA Polymerase [(0.3 U); Invitrogen™].

Samples were amplified in a Dual 96-Well GeneAmp® PCR System 9700 thermocycler (Applied Biosystems™) in the following conditions: initial denaturation at 95°C for 15 min, followed by 45 cycles of 30 s at 95°C, 45 s at temperatures between 52°C and 60°C (variable with the loci) and 30 s at 72°C; a final elongation step was held for 60 min at 72°C.

PCR products were checked in 2% agarose gel and stained with ethidium bromide. According to the quality of the amplification, the samples were diluted in water, mixed with formamide and LIZ® 500-bp internal size standard (Applied Biosystems™) and detected by capillary electrophoresis using a 3100 Genetic Analyzer® (Applied Biosystems™) sequencer.

The software used to visualize the results was GeneMapper® Software v4.0 (Applied Biosystems™).

In order to ensure reliability of our dataset, genotypes for non-invasive samples were achieved after obtaining three consensual scores for heterozygous and four for homozygous individuals, according to a multiple tube methodology (Taberlet *et al.* 1996). Multiplex loading was possible through the use of different colored fluorescent markers for loci that had overlapping size ranges (Table 1).

Table 1 - Characteristics for the 15 polymorphic microsatellite loci - primer sequences, fluorescent label, size range, PCR hybridization temperature (TA) and reference

Locus	Primer sequences (5´- 3´)	Dye Label	Size Range	TA (°C)	Reference
AHT4	AACCGCCTGAGCAAGGAAGT	6FAM	162-182	56	Binns *et al.* (1995)
	GCTCCCAGAGAGTTTACCCT				
AHT5	ACGGACACATCCCTGCCTGC	VIC	130-158	58	Binns *et al.* (1995)
	GCAGGCTAAGGGGGCTCAGC				
CA425	AGCTGCCTCGTTAATTCA	PET	248-254	58	Eggleston-Stott *et al.* (1999)
	CTCATGTCCGCTTGTCTC				
COR20	TCTCTACCGCAAGTGAAACC	NED	170-174	60	Hopman *et al.* (1999)
	CTGAATTGTAGGACATCCCG				
COR58	GGGAAGGACGATGAGTGAC	6FAM	220-234	56	Ruth *et al.* (1999)
	CACCAGGCTAAGTAGCCAAAG				
COR90	GGTTTGTCTCTTTGAGGTGTG	NED	106-120	60	Tallmadge *et al.* (1999)
	TGCTCATATCTTCACCCTGC				
HMS6	GAAGCTGCCAGTATTCAACCATTG	VIC	176-188	56	Guékin *et al.* (1994)
	CTCCATCTTGTGAAGTGTAACTCA				
HMS7	CAGGAAACTCATGTTGATACCATC	6FAM	186-190	56	Guékin *et al.* (1994)
	TGTTGTTGAAACATACCTTGACTGT				
HMS20	TGGGAGAGGTACCTGAAATGTAC	6FAM	132-144	54	Guérin and Bertaud (1996)
	GTTGCTATAAAAAATTGTCTCCCTAC				
HTG6	TAATACGACTCACTATAGG	VIC	100-106	54	Marklund *et al.* (1994)
	GTTCACTGAATGTCAAATTCTGCT				
LEX68	AAATCCGAGCTAAAATGTA	PET	170-180	54	Coogle and Bailey (1999)
	TAGGAAGATAGGATCACAAGG				
LEX74	AAGAGTGCTCCCGTGTG	PET	184-192	56	Bailey *et al.* (2000)
	GACAATGCAGAACTGGGTAA				
NVHEQ18	GGAGGAGACAGTGGCCCCAGTC	VIC	106-118	58	Røed *et al.* (1997)
	GCTGAGCTCTCCCATCCCATCG				
UM11	TGAAAGTAGAAAGGGATGTGG	NED	178-186	56	Meyer *et al.* (1997)
	TCTCAGAGCAGAAGTCCCTG				
VHL20	CAAGTCCTCTTACTTGAAGACTAG	6FAM	96-110	58	van Haeringen *et al.* (1994)
	AACTCAGGGAGAATCTTCCTGAG				

Statistical analysis

GENEPOP software (Raymond & Rousset 1995) was used to test for deviations from Hardy-Weinberg proportions as well as to test for gametic (linkage) disequilibrium for each pair of loci

and calculate Fis values (Weir & Cockerham 1984). Exact P values were estimated from the Markov chain algorithm using 10,000 dememorization steps, 500 batches and 5,000 iterations per batch. The number of alleles per locus, observed and expected heterozygosities, as well as polymorphic information content, probabilities of exclusion and estimated null allele frequencies were calculated by Cervus 3.0 software (Kalinowski *et al.* 2007).

Results and discussion

Fifteen out of the 25 tested loci yielded clear PCR products and were polymorphic. Remaining ten loci were discarded either because they failed to amplify consistently (ASB2, COR70, COR75, HMS3, HTG7, HTG10, TKY321, and UCDEQ457) or because they were monomorphic (ASB23, ASB17).

The observed number of alleles ranged from 3 (HMS7, VHL20, COR20, COR90) to 9 (AHT4, AHT5) and expected heterozigosity (H_E) ranged from 0.833 (AHT5) to 0.257 (VHL20) (Table 2). The mean number of alleles across loci was 5.06. Results showed no linkage between all possible loci pairs (P>0.007) after Bonferroni corrections for multiple tests. Two out of 15 loci were strongly deviated from Hardy–Weinberg proportions (AHT5: H_E=0.833, H_O=0.727, P=0.000 and VHL20: H_E=0.257, H_O=0.286, P=0.000).

Significant heterozygosity deficit (P<0.05) due to nonrandom mating was detected in loci AHT5, HMS6 and COR58. HMS6 was also the only locus to present moderately high estimated frequency of null alleles (r>0.2). Potential causes for microsatellite null alleles include poor primer annealing due to nucleotide sequence divergence and PCR failure due to inconsistent DNA template quality or low template quality, which causes some loci/alleles to fail amplification whereas others amplify with relative ease (Dakin & Avise 2004). In the case of our study, we cannot exclude any of these causes; however, our sample set was mainly composed by fecal samples which are known to frequently yield low quantity/quality DNA extracts (Taberlet *et al.* 1996).

Polymorphic information content calculation allowed the identification of AHT5 (PIC=0.789), AHT4 (PIC=0.656) and LEX74 (PIC=0.653) as the most informative loci and VHL20 (PIC=0.229) and HMS20 (PIC=0.331) as the least informative. Obtained values, per locus, for the probability of exclusion of any given candidate parent (P_{EX1}) and of a candidate parent given the genotype of a known parent of the opposite sex (P_{EX2}) were high enough to allow this set of microsatellites to be used in future parentage analysis studies.

Table 2 - African wild ass population summary statistics. N, number of individuals tested, NA number of alleles, H_E unbiased expected heterozygosity, H_O observed heterozygosity, PIC polymorphic information content, P_{EX1} probability of exclusion (both parents unknown), P_{EX2} probability of exclusion (one parent known), F (null) estimated null allele frequency, Fis (Weir and Cockerham 1984)

	AHT4	AHT5	LEX74	CA425	HMS20	HMS6	COR58	LEX68	HMS7	VHL20	UM11	HTG6	COR20	COR90	NVHEQ18
***E. africanus* (N)**	20	22	21	19	22	22	21	21	22	21	22	22	19	22	22
NA	9	9	5	4	5	6	4	6	3	3	5	4	3	5	5
H_E	0,699	0,833	0,715	0,622	0,359	0,656	0,568	0,682	0,65	0,257	0,571	0,638	0,562	0,642	0,515
H_O	0,65	0,727	0,81	0,684	0,409	0,409	0,381	0,524	0,545	0,286	0,682	0,773	0,474	0,636	0,545
PIC	0,656	0,789	0,653	0,536	0,331	0,591	0,488	0,63	0,564	0,229	0,48	0,574	0,445	0,58	0,473
P_{EX1}	0,3	0,46	0,28	0,19	0,06	0,23	0,16	0,27	0,2	0,03	0,16	0,21	0,15	0,22	0,14
P_{EX2}	0,49	0,63	0,48	0,33	0,19	0,4	0,3	0,45	0,35	0,12	0,29	0,38	0,25	0,39	0,3
F (null)	0,027	0,048	0	0	0	0,225	0,185	0,144	0,079	0	0	0	0,073	0,001	0
F_{is}	0,0714	0.1295*	-0,1352	-0,1038	-0,142	0.3824**	0.3347*	0,2361	0,1642	-0,1163	-0,2	-0,1535	0,1606	0,0084	-0,0611

Significance levels are *(P< 0.05), ** (P< 0.01), and *** (P< 0.001)

Since this is the first work that assessed the African wild ass genetic diversity using molecular markers and because this panel of markers showed reasonable high polymorphism level and negligible values of null alleles for nearly all analyzed loci, we propose this set of 15 microsatellites markers as the standard set to be used in future population genetic studies for the African wild ass. In resume, here we showed a cost efficient example on how to use the publically available genetic data from model species to extract molecular markers to use in population and conservation genetics studies of closely related wild species.

Acknowledgements

This work resulted from a project aiming to develop molecular markers to assess population genetics parameters of the African wild ass. The project was supported by an FCT grant PTDC/BIA-BDE/64111/2006.

References

Bailey E., Skow L., Bernoco D., DelValle A., Scavone M.D.P., Bowling A.T. & Murray J.D. (2000) Equine dinucleotide repeat loci LEX071 through LEX078. *Animal Genetics* **31**, 286-7.

Beja-Pereira A., England P.R., Ferrand N., Jordan S., Bakhiet A.O., Abdalla M.A., Mashkour M., Jordana J., Taberlet P. & Luikart G. (2004a) African origins of the domestic donkey. *Science* **304**, 1781.

Beja-Pereira A., Oliveira R., Alves P.C., Schwartz M.K. & Luikart G. (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. *Mol Ecol Resour* **9**, 1279-301.

Beja-Pereira A., Zeyl E.V.E., Ouragh L., Nagash H., Ferrand N., Taberlet P. & Luikart G. (2004b) Twenty polymorphic microsatellites in two of North Africa's most threatened ungulates: Gazella dorcas and Ammotragus lervia (Bovidae; Artiodactyla). *Molecular Ecology Notes* **4**, 452-5.

Binns M.M., Holmes N.G., Rolliman A. & Scott A.M. (1995) The identification of polymorphicmicrosatellite loci in the horse and their use in thoroughbred parentage testing. *British Veterinary Journal* **151**, 9-15.

Boyd L., W. Z. & King S.R.B. (2008) *Equus ferus przewalskii*. URL http://iucnredlist.org.

Coogle L. & Bailey E. (1999) Equine dinucleotide repeat loci LEX064 through LEX070. *Animal Genetics* **30**, 71-2.

Dakin E.E. & Avise J.C. (2004) Microsatellite null alleles in parentage analysis. *Heredity (Edinb)* **93**, 504-9.

Eggleston-Stott M.L., DelValle A., Bautista M., Dileanis S. & Wictum E. (1999) Twelve equine dinucleotide repeats at microsatellite loci UCDEQ304, UCDEQ380, UCDEQ387, UCDEQ411, UCDEQ439, UCDEQ440, UCDEQ455, UCDEQ457, UCDEQ464, UCDEQ465, UCDEQ482 and UCDEQ497. *Animal Genetics* **30**, 69-70.

Guékin G., Bertaud M. & Amigues Y. (1994) Characterization of seven new horse microsatellites: HMS1, HMS2, HMS3, HMS5, HMS6, HMS7 and HMS8. *Animal Genetics* **25**, 62-.

Guérin G. & Bertaud M. (1996) Characterization of two polymorphic horse microsatellites: HMS15 and HMS20 *Animal Genetics* **27**, 123.

Hopman T.J., Han E.B., Story M.R., Schug M.D., Aquadro C.F., Bowling A.T., Murray J.D., Caetano A.R. & Antczak D.F. (1999) Equine dinucleotide repeat loci COR001-COR020. *Animal Genetics* **30**, 225-6.

Kalinowski S.T., Taper M.L. & Marshall T.C. (2007) Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* **16**, 1099-106.

Marklund S., Ellegren H., Eriksson S., Sandberg K., Andersson L., Marklund S., Ellegren H., Eriksson S., Sandberg K. & Andersson L. (1994) Parentage testing and linkage analysis in the horse using a set of highly polymorphic microsatellites. *Animal Genetics* **25**, 19-23.

Meyer A.H., Valberg S.J., Hillers K.R., Schweitzer J.K. & Mickelson J.R. (1997) Sixteen new polymorphic equine microsatellites *Animal Genetics* **28**, 69-70.

Moehlman P.D. (2002) *Equids: Zebras, Asses and Horses: Status Survey and Conservation Action Plan* IUCN/SSC Equid Specialist Group Gland, Switzerland and Cambridge, United Kingdom.

Moehlman P.D., Yohannes, H., Teclai, R. & Kebede, F. (2008) *Equus africanus*.

Raymond M. & Rousset F. (1995) GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. *Journal of Heredity* **86**, 248-9.

Røed K.H., Midthjell L., Bjørnstad G. & Olsaker I. (1997) Equine dinucleotide repeat microsatellites at the NVHEQ5, NVHEQ7, NVHEQ11, NVHEQ18 and NVHEQ24 loci. *Animal Genetics* **28**, 381-2.

Ruth L.S., Hopman T.J., Schug, Aquadro C.F., Bowling A.T., Murray J.D., Caetano A.R. & Antczak D.F. (1999) Equine dinucleotide repeat loci COR041-COR060. *Animal Genetics* **30**, 320-1.

Taberlet P., Griffin S., Goossens B., Questiau S., Manceau V., Escaravage N., Waits L.P. & Bouvet J. (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Res* **24**, 3189-94.

Tallmadge R.L., Evans K.G., Hopman T.J., Schug M.D., Aquadro C.F., Bowling A.T., Murray J.D., Caetano A.R. & Antczak D.F. (1999) Equine dinucleotide repeat loci COR081–COR100. *Animal Genetics* **30**, 462-78.

van Haeringen H., Bowling A.T., Stott M.L., Lenstra J.A. & Zwaagstra K.A. (1994) A highly polymorphic horse microsatellite locus: VHL20. *Animal Genetics* **25**, 207-.

Weir B.S. & Cockerham C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, 1358-70.

Article 3

In preparation

Non-invasive genetic assessment of critically endangered African wild ass (Equus africanus) populations

Sónia Rosenbom, Fanuel Kebede, Redae Teclai, Hagos Yohannes, Futsum Hagos, Patricia Moehlman and Albano Beja-Pereira

Abstract

During the last century anthropogenic pressures (e.g., hunting and habitat degradation) together with rapid environmental changes have been driving the African wild ass (*Equus africanus*) to the brink of extinction. Although, this species is considered the rarest and most threatened wild equid it is also the less studied one. We have collected African wild ass fecal samples from two locations in Ethiopia and one in Eritrea, in order to assess levels of genetic diversity, population structure and demographic parameters, using genotype data from 10 polymorphic microsatellite markers. Sequence analysis of a fragment of the hypervariable region of the mtDNA was used to identify the matrilineal origin of morphologically identified African wild ass individuals. Obtained results revealed the absence of geographic structuring among extant African wild asses and the Eritrean population as the one possessing the highest values of genetic diversity (H_E=0.63; Na=4.7). F_{ST} among these populations was estimated at 0.10 (P<0.05), confirming a scenario of low population structure. Bidirectional historical migration as well as recent migration were detected between the Ethiopian and Eritrean populations. Effective population sizes for both Ethiopian (*Ne* = 26.2) and Eritrean (*Ne* = 25.6) populations were low, confirming these populations as extremely vulnerable. One individual belonging to the Eritrean population was identified as a first generation hybrid, supporting the existence of rare and geographically limited hybridization between African wild asses and domestic donkeys. This study stands as a first ever genetic assessment of the African wild ass and confirms the high level of endangerment this species is currently facing.

Introduction

Over the past two centuries most of the large African mammals have seen their numbers greatly reduced mainly due to human-mediated habitat destruction, poaching and environmental changes (Cardillo *et al.* 2005). The African wild ass (*Equus africanus*) is no exception to this fact. Once distributed across a wider range, north and west into Sudan and Egypt, its current distribution range is now believed to be circumscribed to the Danakil Desert of Ethiopia and Eritrea (Moehlman *et al.* 2008).

Despite the significant decrease in numbers, demographic trends in Ethiopian and Eritrean populations are somewhat different. In Ethiopia, there has been reports of a severe population decline since the early 1970s (approximately 95%) and the largest remaining population is now located in the north-eastern Afar region. The number of individuals in this country is now believed to be less than 200 with a density of approximately 0.625 animals per 100 square kilometres. Long-term data are not available in Eritrea, but since the mid-1990's the population appears to be stable and in a limited study area the density is approximately 47 individuals per 100 square kilometres (Moehlman *et al.* 2008). Overall number of African wild asses in the wild is estimated in less than 600 animals and the species is currently listed as critically endangered by IUCN (Moehlman *et al.* 2008) and is, by definition, facing an extremely high risk of extinction in the wild.

Both Ethiopian and Eritrean authorities have undergone efforts in order to prevent further loss of individuals, however major threats subsist across the species' distribution range: hunting for food and medicinal purposes, limited access to drinking water and forage (largely due to competition with livestock) and the possible hybridization with the domestic donkey (Moehlman *et al.* 2008). Given the current scenario, it is urgent the design and implementation of efficient conservation programs for the remaining populations of this large wild ungulate, however, before aiming to this goal, it is yet necessary to evaluate the "genetic health" and structure of the remaining subpopulations.

The genetic health of a population is directly dependent from its effective size (*Ne*) (Wright 1931), defined as the size of an ideal population experiencing the same rate of random genetic change over time as the real population under consideration. *Ne* estimation has become an important parameter in conservation and management programs (Schwartz *et al.* 2007), mainly because it can help to assess the population viability and the impact that genetic drift or inbreeding can cause in a particular population. Other important parameters that influence a population's genetic health are those concerning how genetic diversity is distributed within populations and the levels of connectivity between different subpopulations. Assessing levels of gene flow under

metapopulation scenarios is essential in conservation genetics studies of small and endangered species, hence is known that migration could, up to a point, counter the effect of enhanced genetic drift (Palstra & Ruzzante 2008).

Yet another important assessment is on the genetic integrity and potential hybridization between wild and domestic individuals, hence it is known that this process can reduce the fitness of the wild population and cause genetic swamping (Allendorf 2007). As these regions of Africa possess large populations of feral and domestic donkey, this is a potential threat to the integrity of the *Equus africanus* species, as it is the closest wild relative and the most probable ancestral species of the domestic donkey (Beja-Pereira *et al.* 2004; Kimura *et al.* 2011) and, therefore hybridization can occur.

Indeed, in the case of the African wild ass the existence of interbreeding between the wild and the domestic form has been a recurrent theme of concern. Pastoralist communities in the arid regions of the Horn of Africa are known to still use domestic donkeys for transportation of goods and people (Blench 2000), however donkeys are not kept enclosed and are frequently left to wonder around settlements and drinking at the same water pounds as the African wild ass, what makes interaction with roaming wild individuals possible.

Non-invasive sampling have become widely used in conservation genetics studies, because this methodological approach allows the collection of samples from species logistically difficult to sample, without disturbing them, and simultaneously overcomes the ethical issues related to sampling procedures of endangered species (Goossens 2009).

Here, we present the first genetic study on the African wild ass, estimating parameters that will give us an overall idea on the genetic health of extant subpopulations. We have used a non-invasive sampling methodology and sequence analysis of a fragment of the hypervariable region of mtDNA to identify the matrilineal origin of phenotypically wild individuals. In addiction we analysed variation at ten previously tested horse microsatellites (Rosenbom *et al.* 2011) to assess levels of genetic diversity, effective population size, geographic structure and migration among African wild ass populations in Ethiopia and Eritrea. Finally, we investigated the genetic signature of reported bottleneck events in extant African wild ass populations. This study will stand as a reference for the design of future conservation programs and will evaluate, based on genetic data, the current risk of extinction of this species.

Material and Methods

Samples

African wild ass fecal samples were collected in two different locations in Ethiopia (Afdera and Hillu), during monitoring field trips from May 2009 to September 2010 and from one location in Eritrea (Denkelia) during 2012 (February to May) (Fig. 1). Samples were collected on the field and separated in individual paper bags. Their geographical location was recorded as well as any other relevant information, namely morphological features. Samples were then dried naturally and stored at room temperature until further processing. Additionally, three equid skeletons, tentatively identified as African wild asses and collected in Eritrea during field trips, in the years 2006-2007, were added to our data set for both genetic and morphometric analyses.

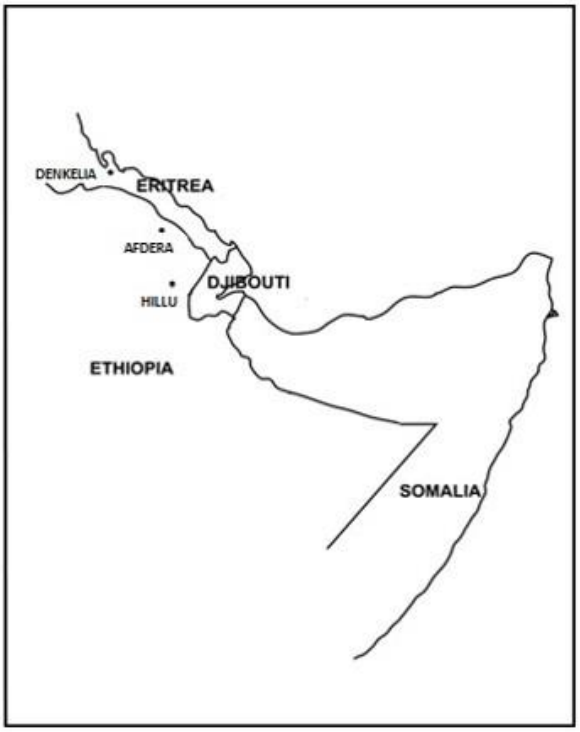


Figure 1 – Sampling locations. Fecal samples were collected once in each indicated location in a period from May 2009 to May 2012

All samples were collected under extreme environmental and climatic conditions (namely extremely high temperatures and high levels of radiation) what has influenced sample quality and limited the number of markers used.

Due to the low number of samples collected in the Ethiopian location of Hillu and given the potential biases introduced by the random capture of some alleles and not others in the calculation of relevant genetic parameters, we have decided to consider only two populations – Ethiopia and Eritrea.

In addiction blood samples from domestic donkeys (N = 26) were collected across the surroundings of the Afar wild ass population, during 2011.

DNA extraction and molecular methods

DNA extractions from fecal samples were carried on a laminar flux chamber, physically separated from the PCR room. Samples were processed in batches with a maximum of 16 samples per set. All material used during the extraction process was sterilized between sample processing. Negative controls were included in each batch of DNA extractions, in order to monitor for possible contaminations. Fecal samples were processed according to an adapted protocol of JETQUICK Tissue DNA Spin Kit (GENOMED) (Costa et al, unpublished manuscript). Blood samples were processed according to the blood sample protocol in the JETQUICK Tissue DNA Spin Kit (GENOMED). Dried tissue around bones was used for DNA extraction in the three skeleton samples. These samples were processed according to the tissue sample protocol in the JETQUICK Tissue DNA Spin Kit (GENOMED), with some modifications, which included prolonged time in the digestion step and an increase in the amount of proteinase K. Samples were then diluted in elution buffer, according to the amount of DNA visible on the gel, and stored at -20ºC.

A 410 bp fragment of the hypervariable region 1 (HVRI) of the mtDNA was amplified using primer pairs and conditions described elsewhere (Rosenbom et al., submitted). Samples that failed to produce PCR amplification of this fragment after three attempts were considered unusable for further genetic analyses.

From a set of 15 horse microsatellites, we have chosen the best ten according to a previously pilot study (Rosenbom *et al.* 2011) in which Mendelian transmission, reproducibility, as well as levels of polymorphism were previously assessed. Forward primers were modified by end labeling with fluorescent dyes (6-FAM™, VIC®, NED™, PET®) at the 5′end. Each 15-µl reaction volume consisted of water, DNA, primers and fluorescent labels (0.06 µM primer forward, 0.6 µM primer reverse and dye), dNTPs (30 mM each), 10× buffer [200 mM Tris–HCl (pH 8.4), 500 mM KCl], BSA (0.4 µg/µl), MgCl2 (variable between 1.5 and 3 mM, according to the loci) and Platinum®

Taq DNA Polymerase ((0.3 U); Invitrogen™). Negative controls were included in all PCR amplifications to monitor potential cross contamination. Samples were amplified in a Dual 96-Well GeneAmp® PCR System 9700 thermocycler (Applied Biosystems™) following conditions described in Rosenbom *et al.* (2011). PCR products were checked in 2% agarose gel stained with GelRed™ and, according to the quality of the amplification, diluted in water, mixed with formamide and LIZ® 500-bp internal size standard (Applied Biosystems™) and detected by capillary electrophoresis using a 3100 Genetic Analyzer® (Applied Biosystems™) sequencer. Software GeneMapper® v4.0 (Applied Biosystems™) was used to score individual genotypes.

Genotyping quality control

In order to ensure reliability of our microsatellite data set we have followed main recommendations outlined by Bonin *et al.* (2004), throughout the entire process, from sample collection to data analysis. This recommendations included: implementation of a standard protocol for collecting, labeling and conserving samples in the field; inclusion of negative controls during lab procedures, use of a dedicated room and lab material for DNA extraction; rigorous protocol for PCR amplification; experience and rigor in genotype scoring. A multiple tubes approach as suggested by Taberlet *et al.* (1996) was implemented, with some modifications. Each sample was initially amplified three times. Loci that were heterozygous in all replicates were scored as reliable, all homozygote and uncertain genotypes were further amplified up to eight times, until four consensual scores were obtained. Samples that could not be reliably typed for at least seven out of the ten amplified microsatellite loci were discarded. We have used software MICRO-CHECKER (Van Oosterhout *et al.* 2004), to detect evidence for scoring errors due to stuttering and large allele dropout (short allele dominance). Finally we used software CERVUS 3.0 (Marshall *et al.* 1998; Kalinowski *et al.* 2007), in order to calculate null allele frequencies and identify matching genotypes in the data set, detecting multiple captures of the same individual.

Statistical analyses

Genetic diversity

Expected (H_E) and observed heterozygosities (H_O), number of alleles (Na) and Wright's F statistics, per locus, were calculated by GenAlEx 6.5 software (Peakall & Smouse 2006; Peakall & Smouse 2012). The same software was used to calculate diversity measures per population (Na, H_O, H_E, uH_E). Deviations from Hardy–Weinberg expectations were tested, using program

GENEPOP 4.2 (Raymond & Rousset 1995), for each population–locus combination. Significance of the departure from Hardy–Weinberg expectations was assessed by exact tests with unbiased P values estimated using a Markov chain method (set to 1000 batches of 10000 iterations each and with 10000 steps of dememorization). A global test across loci and populations was performed using Fisher's method. The same software was used in order to test for gametic (linkage) disequilibrium (LD) between all pairs of loci, across populations.

Population structure

Population structure was estimated with the model-based Bayesian approach implemented in STRUCURE version 2.3.4 (Pritchard *et al.* 2000). Individuals were placed into *K* populations, which corresponded to genetic clusters with distinctive allele frequencies and assigned probabilistically to those populations, with membership coefficients summing up to 1 across clusters. Runs with values of *K* varying from 1 to 5, corresponding to the number of sampled populations plus three, were repeated 10 times, using the admixture model and correlated frequencies. Each run had 10^5 iterations, with the first 10% of iterations discarded as burn-in. In order to achieve a correct estimation of *K*, we have used the statistics Δ*K* introduced by Evanno *et al.* (2005) as implemented in software STRUCTURE HARVESTER (Earl & vonHoldt 2011).

A Principal Coordinate Analysis (PCoA), as implemented in GenAlEx 6.5 (Peakall & Smouse 2006; Peakall & Smouse 2012), was used to visualize genetic differentiation and relationship between African wild ass populations.

Due to high frequency of null allele (r>0.20) at two microsatellite loci we have used software FREENA (Chapuis & Estoup 2007) in order to calculate the F_{ST} estimator Θ (Weir & Cockerham 1984) of population differentiation. This software estimates population differentiation, by correcting the positive bias induced by null alleles and provides 95% confidence intervals on obtained values. In addition, F_{ST} was also calculated using GenAlEx 6.5 (Peakall & Smouse 2006; Peakall & Smouse 2012), in order to test for congruence of obtained results using different estimation methods. Significance of obtained results was assessed by performing 1000 permutations.

Assessing hybridization

To identify the matrilineal origin of phenotypically wild individuals we have used sequence based analyses of a 410 bp fragment of the hypervariable region of the mtDNA. Sequence trace files were edited in DNAStar 6.0 (DNAStar Inc., Madison, WI) and aligned by software Mega version 5.1 (Tamura *et al.* 2011).

In order to assess the power of our markers to identify putative hybrids in the data set the following steps were taken. We have ran software STRUCTURE 2.3.4 (Pritchard *et al.* 2000) for the data set including all wild and domestic individuals (*K*=2) and chose the 20 individuals from each population that presented individual membership proportions (*q*) higher than 0.90 (and small 90% CI) to define "true" parental classes, in order to avoid biases caused by putative hybrids. We have then simulated 40 individuals for each hybrid class (F1, backcrosses to wild and backcrosses to domestic) using software HYBRIDLAD 1.0 (Eg Nielsen *et al.* 2006) and conducted a new STRUCTURE analyses using defined parental and hybrid classes, in order to define the range of *q*-values expected for each simulated hybrid class.

Finally, we have used software STUCTURE 2.3.4 (Pritchard et al. 2000) in order to assign all individuals (wild and domestic) to obtained genetically distinct clusters. The ideal value of *K* was estimated by using Δ*K* statistics introduced by Evanno *et al.* (2005) and implemented in software STRUCTURE HARVESTER (Earl & vonHoldt 2011).

All analyses were performed using the "admixture model" and assuming correlated allele frequencies among populations. Parameters were set at 10^5 iterations following a burn-in period of 10^4 iterations, with ten repetitions for each K value.

A matrix with the genetic distance between individuals, computed from the microsatellite data, was analyzed using a Principal Coordinates Analyses (PCoA), to better visualize the relationships between wild and domestic individuals.

Effective population size estimation and bottleneck detection

To estimate effective population size we have used a "one-sample" estimation method, as implemented in software ONeSAMP 1.1 (Tallmon *et al.* 2008). This software uses summary statistics and approximate Bayesian computation to estimate *Ne* from a single sample of microsatellite data. By using information provided by the user, ONeSAMP creates 50,000 simulated populations, with effective size drawn from a uniform random number between the lower and upper *Ne*, which is set as a prior, and is assumed to come from a population with an initial level of genetic variation determined by theta: the product of historic effective size and the

mutation rate (4 μ). Each simulated population reproduces following a Wright–Fisher model for two to eight generations before being sampled. For each simulated population, ONeSAMP draws samples with identical numbers of individuals and loci to those contained in the target data set. The *Ne* values from simulated populations with summary statistic values close to the summary statistic values from the target population are accepted. Then, the *Ne* values from the accepted simulated populations are used in a weighted local regression to infer the effective size of the target population.

We have used the heterozygosity excess and the mode-shift indicator tests, as implemented in software BOTTLENECK 1.2.02 (Piry *et al.* 1999), in order to detect pronounced reduction in population sizes. This method is based on the expectation that in a bottlenecked population, as allele numbers are reduced faster than gene diversity, the observed heterozygosity will be higher than expected for a population at mutation-drift equilibrium (Cornuet & Luikart 1996). We have used two different microsatellite mutation models, the SMM and a two-phase mutation model (TPM) (Di Rienzo *et al.* 1994), with 95% SMM and 5% IAM, with 12% variance of multiple-step mutations. A total of 5000 simulation iterations were conducted. Statistical significance of the obtained results was assessed with three different tests: the standardized differences test, the one-tailed Wilcoxon's test and the sign test.

Migration

Historic migration between the Ethiopian and the Eritrean populations was assessed by software MIGRATE-N (Beerli 2006). MIGRATE-N estimates historic migration rates between populations using a migration matrix model that includes asymmetric migration rates and different subpopulation sizes, using a Bayesian framework, in order to estimate parameters. We have calculated *M,* the mutation scaled immigration rate that represents the importance of variability brought into the population by immigration compared with the variability created by mutation. In order to ensure congruence of obtained results three independent runs were performed with the following settings: one long chain with a sampling increment of 20 steps and 10000 recorded steps and a burn-in of 20000 steps. Each long chain was replicated 10 times, for each independent run.

Contemporary migration was assess by using the "first generation migrant detection" function implemented in GENECLASS 2.0 software (Piry *et al.* 2004). We have used both Bayesian criteria (Rannala & Mountain 1997; Baudouin & Lebrun 2001) and simulated 10000 individuals according

to the simulation algorithm by Paetkau *et al.* (2004). A Monte-Carlo resampling strategy was used in order to obtain probability values.

Results

Genotyping quality control

No evidence of scoring errors due to stuttering or large allele dropout (short allele dominance) were found across loci, however we found indication of moderate high frequency of null alleles (0.3> r >0.2) at two microsatellite loci. To understand the impact of this moderate frequency of null alleles at these two loci, all posterior analyses were performed including and excluding data from these two loci. Comparative analyses of obtained results showed no differences and so presented results include all ten loci.

Multiple captures of the same individual were detected for individuals captured in the Afdera location in Ethiopia. Individual identity analyses were performed with a threshold of 6 out of 10 matching genotypes, in order to identify possible re-captures. Obtained results revealed that two individuals captured in Afdera, where population density is low and sampling was done without visual identification of individuals, were captured multiple times. Individual AWA68 was captured three additional times and individual AWA78 was re-captured once, with 10 out of 10 and 8 out 10 matching loci, respectively. Duplicated genotypes originated from multiple captures were removed from the final genotype file, used for subsequent analyses.

Genetic diversity

A total of 54 alleles were found across loci, with expected heterozygosity (H_E) ranging from 0.3 to 0.61, and a mean number of alleles across loci of 5.4 (Table 1).

Three out of ten analyzed microsatellite loci strongly deviated from HW expectations, across populations, presenting a highly significant deficit of heterozygous ($P<0.001$) after Bonferroni corrections for multiple tests. Out of the three deviating locus, only locus HMS7 also presented high frequency of null alleles ($r > 0.20$) what could explain observed homozygous excess. Significant non-random association (LD) between loci was not found, and all the loci were treated as statistically independent.

Genetic diversity parameters, namely number of alleles (Na), expected (H_E) and unbiased expected (uH_E) heterozigosities, were higher in the Eritrean population (Table 2).

Table 1 – Summary statistics for 10 microsatellite loci. Allele sample size (k), number of alleles (Na), observed (H_o) and expected (H_E) heterozygosities and Wright´s F statistics

Locus	k	Na	H_O	H_E	Fis	Fit	Fst
AHT4	46	4	0,65 ± 0,09	0,56 ± 0,06	-0,15	-0,07	0,07
AHT5	44	8	0,47 ± 0,04	0,51 ± 0,11	0,07	0,34	0,3
HTG6	48	4	0,59 ± 0,07	0,61 ± 0,05	0,03	0,12	0,09
COR20	51	5	0,42 ± 0,05	0,55 ± 0,04	0,23*	0,37	0,19
COR90	52	5	0,20 ± 0,04	0,3 ± 0,14	0,32	0,44	0,18
HMS7	50	4	0,23 ± 0,02	0,41 ± 0,01	0,43 *	0,56	0,23
HMS20	52	6	0,42 ± 0,17	0,48 ± 0,07	0,13	0,31	0,21
HMS6	49	5	0,32 ± 0,18	0,35 ± 0,19	0,08*	0,21	0,14
UM11	50	7	0,63 ± 0,09	0,59 ± 0,09	-0,06	0,03	0,08
LEX74	46	6	0,4 ± 0,2	0,6 ± 0,01	0,34	0,46	0,19

Table 2 – Summary statistics of sampled African wild ass populations. Represented parameters are mean values per population, across loci. Number of samples is indicated (N) as well as number of different alleles (Na), observed heterozygosity (H_O), expected heterozygosity (H_E), unbiased expected heterozygosity (uH_E) and fixation index (F)

Species	Country	Sampling location	N	Na	H_O	H_E	uH_E	F
Equus africanus somaliensis	Ethiopia	Hillu	5	4.2 ± 0.42	0.48 ± 0.05	0.51 ± 0.05	0.52 ± 0.05	0.04 ± 0.08
		Afdera	29					
	Eritrea	Asmera	19	4.7 ± 0.3	0.44 ± 0.05	0.61 ± 0.03	0.63 ± 0.03	0.29 ± 0.06

Population structure

STRUCTURE (Pritchard et al 2001) analyses of microsatellite genotypes, showed that *K* = 2 produced the highest value of Δ*K* (Fig. 2), however there was no support for geographic structuring of individuals according to their place of origin. Individuals sampled in the Hillu location of Ethiopia, clustered along individuals belonging to the Eritrean population, what could be

justified by the random sampling of very few individuals in this Ethiopian location. PCoA revealed similar results, with the first coordinate justifying 15.5% of the found variation and roughly separating individuals collected in the Afdera region from individuals sampled in the Denkelia and Hillu locations.

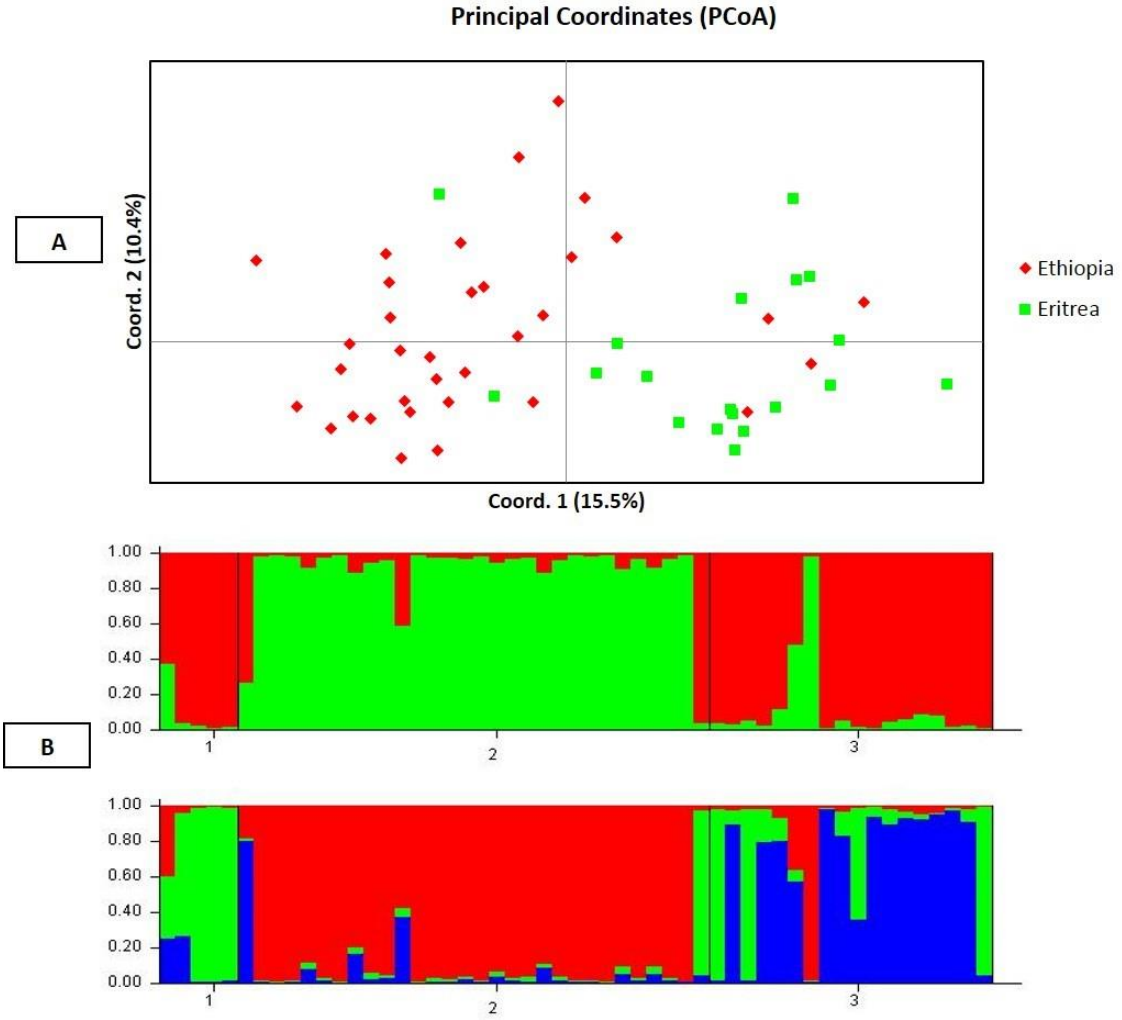


Figure 2 – Principal coordinates analysis showing relationship among multilocus genotypes of African wild ass individuals, belonging to sampled populations. Percentage of variation explained by first and second coordinates are shown in brackets (A). Structure analysis (performed assuming k=2 [A] and k=3 [B] distinct genetic clusters) of African wild ass multilocus microsatellite genotypes (B).

Obtained F_{ST} value between the Ethiopian and the Eritrean populations, after the ENA correction described in Chapuis and Estoup (2007) varied between 0.05 and 0.17 (95% CI). Obtained value for the uncorrected pairwise F_{ST} among the populations of Ethiopia and Eritrea was 0.10 ($P<0.05$).

Assessing hybridization

Two out of the thirty-nine obtained mtDNA sequences for the Ethiopian population of Afdera presented a previously published domestic donkey haplotype. A similar scenario was found is the Eritrean population of Denkelia, with twelve out of the twenty-four obtained sequences being identified as belonging to a different domestic donkey haplotype.

We were able to genotype seven out of the fourteen individuals that presented previously reported mtDNA domestic haplotypes. PCoA grouped these individuals within the wild cluster (Fig. 3). Assignment analyses also supported these results, with all seven individuals being assigned to the morphologically defined population of origin (wild).

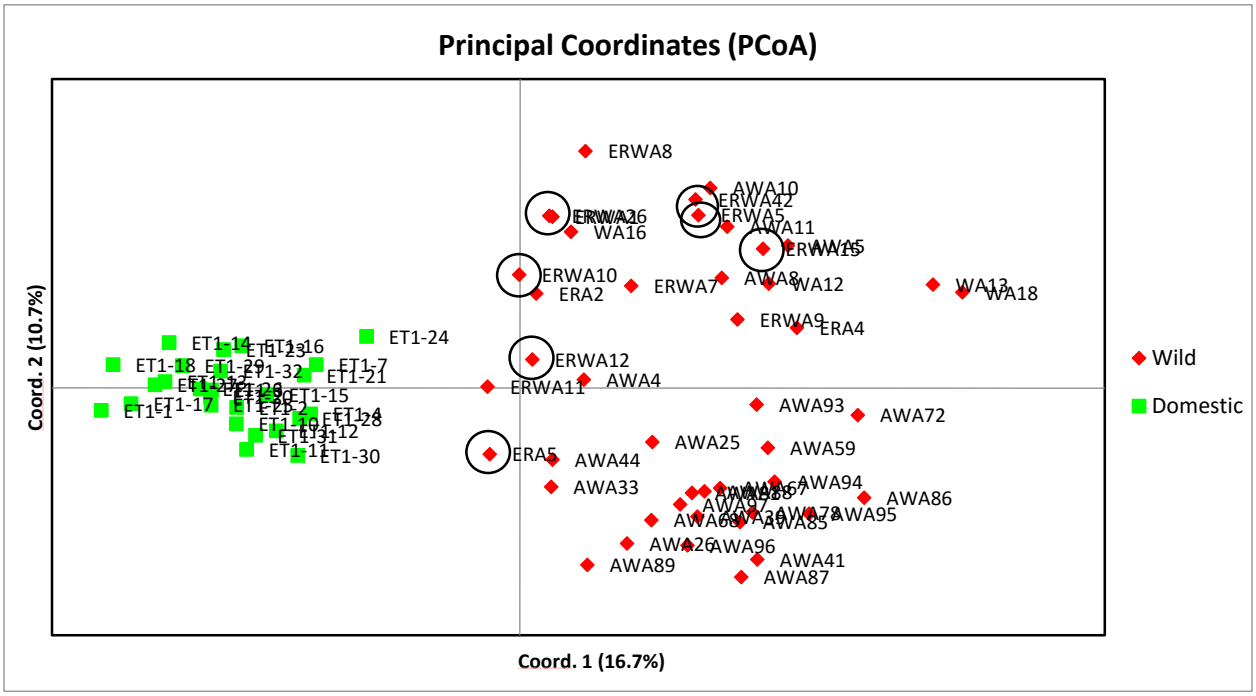


Figure 3 – Principal coordinates analysis showing relationship among multilocus genotypes of African wild ass individuals and domestic donkeys, belonging to the Afar region. Percentage of variation explained by first and second coordinates are shown in brackets. Individuals with previously reported domestic donkey mtDNA haplotypes are identified by black circles.

Assessment on the power of our marker set to discriminate hybrid classes showed that at a probabilistic threshold of $q_i > 0.80$, the admixture analysis performed on simulated genotypes was able to efficiently recognize 100% of the parental individuals with high average proportion of membership to each defined cluster (Q_{wild} = 0.88 and $Q_{domestic}$= 0.90). All F1 simulated hybrids

were correctly identified as admixed, however 42.5% of simulated backcrossed with domestic and 35% of simulated backcrossed with wild showed a $q_i > 0.80$ to a single cluster and could not be distinguished from parental populations.

STRUCTURE analyses revealed that *K*=3 produced the highest value of Δ*K.* Assignment analyses for *K*=3 showed that all domestic donkeys were correctly assigned to the domestic cluster with high membership proportion values ($Q_{domestic} \geq 0.97$). African wild asses belonging to the Ethiopian were all assigned either to the Ethiopian or to the Eritrean wild ass populations with high *q*-values (supplementary material, Table S1), however 4 individuals belonging to the Eritrean wild ass population showed variable admixture levels, with $q_{domestic}$ varying from 0.14 to 0.33 (supplementary material, Table S1). One individual in particular (ERA5) showed an admixture pattern similar to simulated F1 hybrids ($q_{Ethiopia\_wild} = 0.43$; $q_{Eritrea\_wild} = 0.08$; $q_{domestic} = 0.49$).

Effective population size estimation and bottleneck detection

Effective population size (*Ne*) estimation showed comparable numbers for the Eritrean and Ethiopian populations, with values varying between 25.6 (19.4 - 41.5; 95% CI), for the Eritrean population and 26.2 (19.7-37.6; 95% CI) for the Ethiopian population. Overall *Ne* results seem to be in relative good agreement with the reported census number of less than 600 individuals for the species in Eritrea and Ethiopia (Moehlman *et al.* 2008), given the reference value for the *Ne*/*N* ratio of approximately 0.11 (Frankham 2002; Frankham 2009).

Bottleneck results failed to detect a significant population reduction in both subpopulations, despite reports of a recent and accentuated decline in number of individuals, namely in Ethiopia. Results for both the sign test and the standardized differences test showed no significant values (P>0.05) for heterozygosity excess, for the two hypothesized mutation models (TPM and SMM). Mode-shift tests for both subpopulations showed a normal L-shaped distribution of allelic frequencies as expected in non-bottlenecked populations.

Migration

Analyses of historic migratory patterns among studied populations revealed the existence of both historical and contemporary bilateral migration among studied populations. Assessment of historical migration using the *M* parameter revealed the existence of bilateral migration between the populations of Ethiopia and Eritrea, more intense from Ethiopia to Eritrea ($M_{Ethiopia \rightarrow Eritrea} = 49.0$)

(22.6 – 72.6; 97.5% CI) than from Eritrea to Ethiopia ($M_{Eritrea \rightarrow Ethiopia}$ = 35.0) (12.0 – 56.0: 97.5% CI). Recent migration was also detected, with two individuals belonging to the Ethiopian population being identified as first generation migrants from Eritrea (P<0.01) and one individual belonging to the Eritrean population identified as a first generation Ethiopian migrant. Found migratory patterns among African wild ass populations in Ethiopia and Eritrea support long time connectivity among these populations.

Discussion

Our results revealed the Eritrean population as the one possessing higher values for genetic diversity parameters (Na, H_E and H_O). This population is believed to be the largest remaining one, hence there has been reports of a severe population reduction in Ethiopia, over the past 40 years, with a reported 95% decline in African wild ass numbers (Moehlman *et al.* 2008). Perhaps ongoing conservation actions and the general attitude of local pastoralists towards wildlife in Eritrea, has prevented a similar scenario to what is currently found in Ethiopia, resulting in a more stable and genetically healthier population.

Obtained results shows that genetic differentiation among African wild asses is weak (F_{ST} = 0.10), with no clear geographic structuring among Ethiopian and Eritrean populations. Similar results have been found by analyzing mtDNA control region sequences belonging to African wild asses from Ethiopia and Eritrea (Kimura *et al.* 2011; Rosenbom 2014), with three out of the four obtained mtDNA haplotypes being shared among populations. These results are further supported by the detection of both historic and recent migration between the Ethiopian and the Eritrean populations, what would contribute for the overall lack of geographic structuring among extant African wild ass populations.

Despite reports of a significant reduction on African wild ass numbers, both in Ethiopia and in Eritrea, we were unable to obtain significant results for test statistics in genetic bottleneck analyses. Such fact might be a consequence of the limited power of genetic bottleneck tests when using small sample sizes and/or limited number of loci (Peery *et al.* 2012), or be a consequence of historical low diversity among African wild ass populations.

Long time survival of the African wild ass is also directly related to the potential of this species to overcome the enhanced effects of genetic drift and inbreeding. Obtained results for the effective population sizes showed comparable values for the populations of Eritrea and Ethiopia, despite sampling effort being substantially larger in Ethiopia. Both Ethiopian and Eritrean populations

show concerning effective population sizes (*Ne* = 26.2 and *Ne* = 25.6, respectively) and are facing a high risk of extinction, due to enhanced effects of inbreeding and genetic drift. Overall estimates of effective population sizes for studied populations seem to be in relative good agreement with proposed census number of less than 600 individuals in Eritrea and Ethiopia (Moehlman *et al.* 2008), given the empirical reference value for the *Ne/N* ratio of approximately 0.11 (Frankham 2002, 2003; Frankham 2009).

Interbreeding between the African wild ass and the domestic form has long been seen as a potential threat for the survival of this species. The existence of domestic mtDNA haplotypes, among morphologically identified African wild asses, had already been documented in a previous study (Kimura *et al.* 2011), however, up to now, nuclear data has never been used in order to detect hybridization.

Obtained results have unraveled a complex scenario. The Ethiopian population showed high genetic integrity, with all individuals being clearly assigned to the wild cluster with high membership coefficients and no potential hybridization events being detected. Four individuals belonging to the Eritrean population showed variable levels of admixed ancestry and one in particular (ERA5), showed an admixed ancestry pattern in line with simulated F1 hybrids. This individuals was identified, *a posteriori*, as being one of the three collected skeletons in Eritrea, which have undergone morphometric analyses. Obtained information from morphometric data, has identified this individual as an old male, small in size, however possessing large canines. Based on morphometric analyses the individual was identified as a domestic donkey, what supports the hypothesis of an admixed ancestry, when adding the findings of the genetic analyses.

From the remaining three, only one other individual (ERWA10) showed a membership coefficient to the domestic clade higher than 0.2 (q_{domestic} = 0.33), indicative of a probable admixed ancestry, however and given the limited power of our marker set to discriminate hybrid classes beyond first generation hybrids (F1), we cannot draw any definite conclusions about the ancestry of this individual.

Based on our results, we can confirm that sporadic hybridization between the domestic and the wild form was detected in the Eritrean population.

Questions regarding the existence and/or extent of hybridization between the African wild ass and domestic donkeys have long been a theme of debate. In some regions of Northern Africa, herders are known to have attempted breeding their female donkeys with male wild asses, in order to

obtain stronger foals (Baker 1867; Murray 1935; Nicolaisen 1963). If these attempts were successful, the impact of hybridization in the integrity of the African wild ass genetic pool would be minimal, and signs of hybridization should be detected in local domestic donkeys. On the other hand, and since traditionally herders in this geographic region do not enclose their donkeys, it is very likely that throughout times feral individuals would get in touch with wild herds. Whether feral donkeys would survive in wild herds is difficult to assess. In particular, feral female donkeys, roaming in areas controlled by territorial wild males might have a better chance at surviving by joining herds mainly composed by wild females and their offspring. In this case, hybridization could occur and have some impact on the integrity of the African wild ass's genetic pool.

Our study has detected the presence of previously reported domestic mtDNA among African wild asses. However, this fact might not indicative of female driven hybridization, hence only two out of the seven individuals possessing previously reported domestic mtDNA haplotype, showed signs of admixed ancestry, being that the remaining five clustered well among wild individuals, with high individual membership proportion to the wild cluster ($0.83 < q_i < 0.99$).

Based on our data we believe that the detected "domestic" haplotypes among African wild ass are in fact ancestral haplotypes found in pre-domestication African wild ass populations. Those haplotypes would have be retained in contemporary African wild ass populations, as well as be captured in domestic donkey populations and increase in frequency as domestic populations expanded. This believe is further supported by the fact that both retrieved haplotypes correspond to wide spread and high frequency haplotypes among domestic donkeys and also by the fact that they are an exact match to those found by Kimura *et al.* (2011) among historical and museum samples of the African wild ass.

Conclusion

Long time survival of the African wild ass requires immediate actions in order to prevent further loss of this species' evolutionary potential. Management and conservation actions will require the joint efforts of wildlife authorities in Ethiopia and Eritrea, in order to prevent the future isolation of extant African wild ass populations in these countries. Potential hybridization between wild and domestic individuals was detected in the Eritrean population, however further analyses including more markers and more samples from the populations in Ethiopia and Eritrea are needed in order to further understand the extent of this events. The extremely low effective population size values found for both populations highlight the vulnerability of this species.

Acknowledgements

This work was funded by FEDER funds though a project co-funded by the EU program COMPETE and the national funds from the Portuguese Foundation for Science and Technology (FCT) (PTDC/BIA-BIC/118107/2010 and FCOMP-01-0124-FEDER- 019757) and from another FCT project (PTDC/BIA-BDE/64111/2006). AB-P is supported by FCT through an IF-FCT contract. PDM was supported by EcoHealth Alliance, Basel Zoo, Saint Louis Zoo and Liberec Zoo.

References

Allendorf F.W., Luikart, G. (2007) *Conservation and the Genetics of Populations* Blackwell Publishing Massachusetts, USA.

Baker S. (1867) *The Nile tributaries of Abyssinia*. Macmillan, London, UK.

Baudouin L. & Lebrun P. (2001) An Operational Bayesian Approach for the Identification of Sexually Reproduced Cross-Fertilized Populations using Molecular Markers. In: *International Symposium on Molecular Markers* (ed. by Doré DB), pp. 81-94. Acta Hort.

Beerli P. (2006) Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* **22**, 341-5.

Beja-Pereira A., England P.R., Ferrand N., Jordan S., Bakhiet A.O., Abdalla M.A., Mashkour M., Jordana J., Taberlet P. & Luikart G. (2004) African origins of the domestic donkey. *Science* **304**, 1781.

Blench R. (2000) The history and spread of donkeys in Africa. In: *Donkeys, people and development* (ed. by Starkey PaF, D.), p. 244. Animal Traction Network for Eastern and Southern Africa (ATNESA), Wageningen, The Netherlands.

Bonin A., Bellemain E., Bronken Eidesen P., Pompanon F., Brochmann C. & Taberlet P. (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* **13**, 3261-73.

Cardillo M., Mace G.M., Jones K.E., Bielby J., Bininda-Emonds O.R., Sechrest W., Orme C.D. & Purvis A. (2005) Multiple causes of high extinction risk in large mammal species. *Science* **309**, 1239-41.

Chapuis M.P. & Estoup A. (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* **24**, 621-31.

Cornuet J.M. & Luikart G. (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**, 2001-14.

Di Rienzo A., Peterson A.C., Garza J.C., Valdes A.M., Slatkin M. & Freimer N.B. (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences USA* **91**, 3166-70.

Earl D.A. & vonHoldt B.M. (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359-61.

Eg Nielsen E., Bach L.A., Kotlicki P., Eg Nielsen E., Bach L.A. & Kotlicki P. (2006) HYBRIDLAB (version 1.0): a program for generating simulated hybrids from population samples. *Mol Ecol Resour* **6**, 971-3.

Evanno G., Regnaut S. & Goudet J. (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14**, 2611-20.

Frankham R. (2003) Genetics and conservation biology. *Comptes Rendus Biologie* **326 Suppl 1**, S22-9.

Frankham R. (2009) Effective population size/adult population size ratios in wildlife: a review. *Genetical Research* **66**, 95.

Frankham R., Ballou, J.D. and Briscoe, D.A. (2002) *Introduction to Conservation Genetics*. University Press, Cambridge, UK.

Goossens B.a.B., M. W (2009) Non-invasive genetic analysis in conservation. In: *Population Genetics for Animal Conservation* (ed. by Bertorelle G), pp. 167-201. Cambridge University Press, Cambridge, UK

Kalinowski S.T., Taper M.L. & Marshall T.C. (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* **16**, 1099-106.

Kimura B., Marshall F.B., Chen S., Rosenbom S., Moehlman P.D., Tuross N., Sabin R.C., Peters J., Barich B., Yohannes H., Kebede F., Teclai R., Beja-Pereira A. & Mulligan C.J. (2011) Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication. *Proceedings of the Royal Society B: Biological Sciences* **278**, 50-7.

Marshall T.C., Slate J., Kruuk L.E. & Pemberton J.M. (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* **7**, 639-55.

Moehlman P.D., Yohannes H., Teclai R. & Kebede F. (2008) *Equus africanus*. URL http://www.iucnredlist.org.

Murray G.W. (1935) *Sons of Ishmael*. Routledge, London, UK.

Nicolaisen J. (1963) *Ecology and culture of the Pastoral Tuareg*. The National Musum of Copenhagen, Copenhagen, Denmark.

Paetkau D., Slade R., Burden M. & Estoup A. (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* **13**, 55-65.

Palstra F.P. & Ruzzante D.E. (2008) Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Molecular Ecology* **17**, 3428-47.

Peakall R. & Smouse P.E. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics* **28**, 2537-9.

Peakall R.O.D. & Smouse P.E. (2006) genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-95.

Peery M.Z., Kirby R., Reid B.N., Stoelting R., Doucet-Beer E., Robinson S., Vasquez-Carrillo C., Pauli J.N. & Palsboll P.J. (2012) Reliability of genetic bottleneck tests for detecting recent population declines. *Molecular Ecology* **21**, 3403-18.

Piry S., Alapetite A., Cornuet J.M., Paetkau D., Baudouin L. & Estoup A. (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity* **95**, 536-9.

Piry S., Luikart G. & Cornuet J.-M. (1999) Computer note. BOTTLENECK: a computer program for detecting recent reductions in the effective size using allele frequency data. *Journal of Heredity* **90**, 502-3.

Pritchard J.K., Stephens M. & Donnelly P. (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155**, 945-59.

Rannala B. & Mountain J.L. (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences* **94**, 9197-201.

Raymond M. & Rousset F. (1995) GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. *Journal of Heredity* **86**, 248-9.

Rosenbom S., Costa V., Steck B., Moehlman P. & Beja-Pereira A. (2011) Cross-species genetic markers: a useful tool to study the world's most threatened wild equid—*Equus africanus*. *European Journal of Wildlife Research* **58**, 609-13.

Schwartz M.K., Luikart G. & Waples R.S. (2007) Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology and Evolution* **22**, 25-33.

Taberlet P., Griffin S., Goossens B., Questiau S., Manceau V., Escaravage N., Waits L.P. & Bouvet J. (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* **24**, 3189-94.

Tallmon D.A., Koyuk A., Luikart G. & Beaumont M.A. (2008) COMPUTER PROGRAMS: onesamp: a program to estimate effective population size using approximate Bayesian computation. *Molecular Ecology Resources* **8**, 299-301.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M. & Kumar S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**, 2731-9.

Van Oosterhout C., Hutchinson W.F., Wills D.P.M. & Shipley P. (2004) micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-8.

Weir B.S. & Cockerham C.C. (1984) Estimating F-statistics for the analysis of popualtion structure. *Evolution* **38**, 1358-70.

Wright S. (1931) Evolution in Mendelian Populations. *Genetics* **16**, 97-159.

Supplementary material

Table S1 - Individual membership proportions of sampled individuals belonging to African wild ass populations (Ethiopia and Eritrea) and a domestic donkey population, to each obtained cluster (Ethiopia_wild, Eritrea_wild, Domestic)

Sample code	Population	$q_{\text{Ethopia\_wild}}$	$q_{\text{Eritrea\_wild}}$	$q_{domestic}$
AWA4	Ethiopia	0.287 (0.000, 0.873)	0.690 (0.046, 1.000)	0.023 (0.000, 0.166)
AWA5	Ethiopia	0.016 (0.000, 0.102)	0.979 (0.869, 1.000)	0.005 (0.000, 0.028)
AWA8	Ethiopia	0.015 (0.000, 0.093)	0.977 (0.855, 1.000)	0.008 (0.000, 0,044)
AWA10	Ethiopia	0.005 (0.000, 0.026)	0.990 (0.941, 1.000)	0.005 (0.000, 0.026)
AWA11	Ethiopia	0.009 (0.000, 0.049)	0.985 (0.908, 1.000)	0.006 (0.000, 0.034)
AWA25	Ethiopia	0.985 (0.907, 1.000)	0.008 (0.000, 0.045)	0.007 (0.000, 0.038)
AWA26	Ethiopia	0.986 (0.916,1.000)	0.006 (0.000,0.031)	0.008 (0.000,0.044)
AWA33	Ethiopia	0.967 (0.796,1.000)	0.016 (0.000,0.085)	0.017 (0.000,0.112)
AWA39	Ethiopia	0.951 (0.725,1.000)	0.041 (0.000,0.258)	0.008 (0.000,0.047)
AWA41	Ethiopia	0.966 (0.806,1.000)	0.026 (0.000,0.168)	0.009 (0.000,0.056)
AWA44	Ethiopia	0.980 (0.879,1.000)	0.007 (0.000,0.038)	0.012 (0.000,0.074)
AWA59	Ethiopia	0.984(0.908,1.000)	0.011 (0.000,0.055)	0.005 (0.000,0.028)
AWA67	Ethiopia	0.979 (0.867,1.000)	0.015 (0.000,0.098)	0.006 (0.000,0.034)
AWA68	Ethiopia	0.983 (0.895,1.000)	0.008 (0.000,0.046)	0.009 (0.000,0.050)
AWA72	Ethiopia	0.966 (0.795,1.000)	0.029 (0.000,0.185)	0.005 (0.000,0.029)
AWA78	Ethiopia	0.979 (0.870,1.000)	0.014 (0.000,0.095)	0.006 (0.000,0.035)
AWA81	Ethiopia	0.981 (0.884,1.000)	0.012 (0.000,0.076)	0.006 (0.000,0.034)
AWA85	Ethiopia	0.936 (0.688,1.000)	0.022 (0.000,0.149)	0.042 (0.000,0.244)
AWA86	Ethiopia	0.979 (0.864,1.000)	0.016 (0.000,0.113)	0.005 (0.000,0.026)
AWA87	Ethiopia	0.988 (0.926,1.000)	0.005 (0.000,0.025)	0.007 (0.000,0.043)
AWA88	Ethiopia	0.988 (0.929,1.000)	0.007 (0.000,0.035)	0.005 (0.000,0.027)
AWA89	Ethiopia	0.985 (0.907,1.000)	0.005 (0.000,0.028)	0.010 (0.000,0.058)
AWA93	Ethiopia	0.832 (0.406,1.000)	0.155 (0.000,0.578)	0.013 (0.000,0.087)
AWA94	Ethiopia	0.981 (0.882,1.000)	0.011 (0.000,0.063)	0.008 (0.000,0.051)
AWA95	Ethiopia	0.955 (0.752,1.000)	0.033 (0.000,0.221)	0.012 (0.000,0.085)
AWA96	Ethiopia	0.981 (0.877,1.000)	0.011 (0.000,0.067)	0.008 (0.000,0.049)
AWA97	Ethiopia	0.985 (0.909,1.000)	0.005 (0.000,0.029)	0.009 (0.000,0.056)
WA12	Eritrea	0.023 (0.000,0.165)	0.970 (0.808,1.000)	0.007 (0.000,0.038)
WA13	Eritrea	0.028 (0.000,0.205)	0.968 (0.785,1.000)	0.004 (0.000,0.023)
WA16	Eritrea	0.018 (0.000,0.129)	0.930 (0.600,1.000)	0.052 (0.000,0.359)
WA18	Eritrea	0.037 (0.000,0.269)	0.959 (0.725,1.000)	0.004 (0.000,0.021)
ERA2	Eritrea	0.018 (0.000,0.105)	0.947 (0.649,1.000)	0.035 (0.000,0.271)
ERA4	Eritrea	0.061 (0.000,0.414)	0.930 (0.575,1.000)	0.009 (0.000,0.057)
ERA5	Eritrea	0.429 (0.000,0.981)	0.084 (0.000,0.484)	0.487 (0.000,0.989)

ERWA1	Eritrea	0.016 (0.000,0.109)	0.959 (0.741,1.000)	0.025 (0.000,0.169)
ERWA5	Eritrea	0.065 (0.000,0.349)	0.921 (0.626,1.000)	0.015 (0.000,0.098)
ERWA7	Eritrea	0.009 (0.000,0.051)	0.963 (0.772,1.000)	0.027 (0.000,0.192)
ERWA8	Eritrea	0.007 (0.000,0.039)	0.977 (0.855,1.000)	0.016 (0.000,0.112)
ERWA9	Eritrea	0.022 (0.000,0.145)	0.968 (0.799,1.000)	0.010 (0.000,0.062)
ERWA10	Eritrea	0.118 (0.000,0.563)	0.551 (0.000,1.000)	0.330 (0.000,0.943)
ERWA11	Eritrea	0.084 (0.000,0.536)	0.772 (0.029,1.000)	0.145 (0.000,0.593)
ERWA12	Eritrea	0.103 (0.000,0.623)	0.723 (0.000,1.000)	0.174 (0.000,0.627)
ERWA15	Eritrea	0.011 (0.000,0.066)	0.980 (0.875,1.000)	0.009 (0.000,0.052)
ERWA26	Eritrea	0.014 (0.000,0.088)	0.942 (0.662,1.000)	0.045 (0.000,0.305)
ERWA42	Eritrea	0.006 (0.000,0.031)	0.987 (0.919,1.000)	0.008 (0.000,0.040)
ET1-1	Domestic	0.004 (0.000,0.023)	0.004 (0.000,0.019)	0.992 (0.952,1.000)
ET1-2	Domestic	0.020 (0.000,0.145)	0.010 (0.000,0.062)	0.970 (0.810, 1.000)
ET1-3	Domestic	0.021 (0.000,0.143)	0.016 (0.000,0.109)	0.963 (0.797,1.000)
ET1-4	Domestic	0.087 (0.000,0.453)	0.016 (0.000,0.113)	0.897 (0.529,1.000)
ET1-7	Domestic	0.009 (0.000,0.051)	0.022 (0.000,0.151)	0.970 (0.806,1.000)
ET1-10	Domestic	0.044 (0.000,0.295)	0.015 (0.000,0.099)	0.942 (0.674,1.000)
ET1-11	Domestic	0.013 (0.000,0.084)	0.007 (0.000,0.038)	0.980 (0.874,1.000)
ET1-12	Domestic	0.010 (0.000,0.057)	0.009 (0.000,0.056)	0.980 (0.878,1.000)
ET1-13	Domestic	0.005 (0.000,0.030)	0.006 (0.000,0.032)	0.989 (0.931,1.000)
ET1-14	Domestic	0.004 (0.000,0.024)	0.008 (0.000,0.046)	0.988 (0.924,1.000)
ET1-15	Domestic	0.008 (0.000,0.045)	0.009 (0.000,0.054)	0.983 (0.891,1.000)
ET1-16	Domestic	0.004 (0.000,0.022)	0.007 (0.000,0.037)	0.989 (0.935,1.000)
ET1-17	Domestic	0.005 (0.000,0.029)	0.006 (0.000,0.033)	0.988 (0.930,1.000)
ET1-18	Domestic	0.004 (0.000,0.023)	0.005 (0.000,0.028)	0.991 (0.943,1.000)
ET1-20	Domestic	0.011 (0.000,0.068)	0.015 (0.000,0.101)	0.974 (0.838,1.000)
ET1-21	Domestic	0.023 (0.000,0.162)	0.013 (0.000,0.089)	0.964 (0.794,1.000)
ET1-23	Domestic	0.005 (0.000,0.028)	0.006 (0.000,0.031)	0.989 (0.934,1.000)
ET1-24	Domestic	0.068 (0.000,0.408)	0.065 (0.000,0.338)	0.866 (0.522,1.000)
ET1-25	Domestic	0.008 (0.000,0.047)	0.007 (0.000,0.041)	0.985 (0.904,1.000)
ET1-26	Domestic	0.006 (0.000,0.030)	0.016 (0.000,0.107)	0.978 (0.864,1.000)
ET1-27	Domestic	0.007 (0.000,0.041)	0.005 (0.000,0.026)	0.988 (0.926,1.000)
ET1-28	Domestic	0.014 (0.000,0.088)	0.018 (0.000,0.116)	0.967 (0.796,1.000)
ET1-29	Domestic	0.005 (0.000,0.030)	0.006 (0.000,0.032)	0.989 (0.930,1.000)
ET1-30	Domestic	0.037 (0.000,0.262)	0.004 (0.000,0.023)	0.958 (0.731,1.000)
ET1-31	Domestic	0.013 (0.000,0.079)	0.016 (0.00,0.111)	0.971 (0.814,1.000)
ET1-32	Domestic	0.008 (0.000,0.048)	0.007 (0.000,0.040)	0.985 (0.905,1.000)

Chapter IV

Insights into donkey domestication and ancestry: patterns of past and current variation

Article 4

Proceedings of the Royal Society B, 278 (1702), 50-57

Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication

Birgitta Kimura, Fiona B. Marshall, Shanyuan Chen, **Sónia Rosenbom**, Patricia D. Moehlman, Noreen Tuross, Richard C. Sabin, Joris Peters, Barbara Barich, Hagos Yohannes, Fanuel Kebede, Redae Teclai, Albano Beja-Pereira and Connie J. Mulligan

Abstract

Genetic data from extant donkeys (*Equus asinus*) have revealed two distinct mitochondrial DNA haplogroups, suggestive of two separate domestication events in northeast Africa about 5000 years ago.

Without distinct phylogeographic structure in domestic donkey haplogroups and with little information on the genetic makeup of the ancestral African wild ass, however, it has been difficult to identify wild ancestors and geographical origins for the domestic mitochondrial clades.

Our analysis of ancient archaeological and historic museum samples provides the first genetic information on the historic Nubian wild ass (*Equus africanus africanus*), Somali wild ass (*Equus africanus somaliensis*) and ancient donkey. The results demonstrate that the Nubian wild ass was an ancestor of the first donkey haplogroup. In contrast, the Somali wild ass has considerable mitochondrial divergence from the Nubian wild ass and domestic donkeys. These findings resolve the long-standing issue of the role of the Nubian wild ass in the domestication of the donkey, but raise new questions regarding the second ancestor for the donkey. Our results illustrate the complexity of animal domestication, and have conservation implications for critically endangered Nubian and Somali wild ass.

Introduction

Domestication of the donkey (*Equus asinus*) approximately 5000 years ago transformed ancient societies and land-based transport in Africa and Eurasia, allowing the development of mobile pastoralism and ancient overland trade routes and contributing to the growth of the early Egyptian State (Hassan 1993; Marshall 2007; Rossel *et al.* 2008). Today donkeys are essential means of transport for people living in many mountainous, desert and poor regions of the world (Starkey 2000; Smith & Pearson 2005). Little is known, however, about domestication of the donkey. Historically it has been thought that ancient Egyptians domesticated the African wild ass (*Equus africanus*) although near-eastern domestication has also been suggested (Uerpmann 1991; Clutton-Brock 1992; Vila 2002). Recent research shows the importance of load-bearing donkeys to the earliest pharaohs and emphasizes the slow nature of their morphological, and probably genetic, change during domestication (Rossel *et al.* 2008). Research on the genetics of modern donkeys worldwide demonstrated the existence of two distinct mitochondrial haplogroups, termed Clades 1 and 2 (Beja-Pereira *et al.* 2004; Chen *et al.* 2006; Vilà *et al.* 2006). Genetic variability in both domestic maternal lineages was greatest in Africa, therefore mitochondrial results did not provide support for the hypothesis of Asian domestication. Specifically, Beja-Pereira *et al.* (2004) argued that Asiatic wild asses were excluded as progenitors of modern donkeys and proposed two separate domestication events that occurred on the African continent (Beja-Pereira *et al.* 2004). There are at least three possible African candidates for wild ancestors of the donkey, the Atlas, Nubian and Somali wild ass (Fig. 1), reflecting the fact that distinct geographical patterning does not exist in modern donkey haplogroups.

Based on the available genetic data, it has been hypothesized that Nubian wild asses were the ancestors of donkeys of Clade 1 and that a relative of the Somali wild ass, probably extinct, was the ancestor of Clade 2 (Beja-Pereira *et al.* 2004; Marshall 2007). Archaeological data, the distribution of African wild ass, and linguistic data suggest that mobile African cattle herders domesticated the donkey in response to increasing aridity in the Sahara and the Horn (Beja-Pereira *et al.* 2004; Marshall 2007).

In order to investigate the relationships of African wild ancestors to domestic donkey clades, additional information is needed on variability within and among ancient and modern wild ass populations (Vilà *et al.* 2006).

African wild asses have been well documented in at least three regions of Africa, but there has been debate over the extent to which populations represent the remnants of once continuous variability versus distinct subspecies (Groves 1986; Yalden *et al.* 1986; Moehlman 2002). Nubian

wild asses (*Equus africanus africanus*) were still fairly common in the Atbara region and the Red Sea Hills (NW Sudan) during the first half of the twentieth century AD and the Somali wild ass (*Equus africanus somaliensis*) existed in southern Eritrea, Ethiopia and Somalia (Fig. 1). The Atlas wild ass (*Equus africanus 'atlanticus'*) was once confined to the northwestern part of the continent and probably became extinct in early historic times (Groves 1986) (Fig. 1).

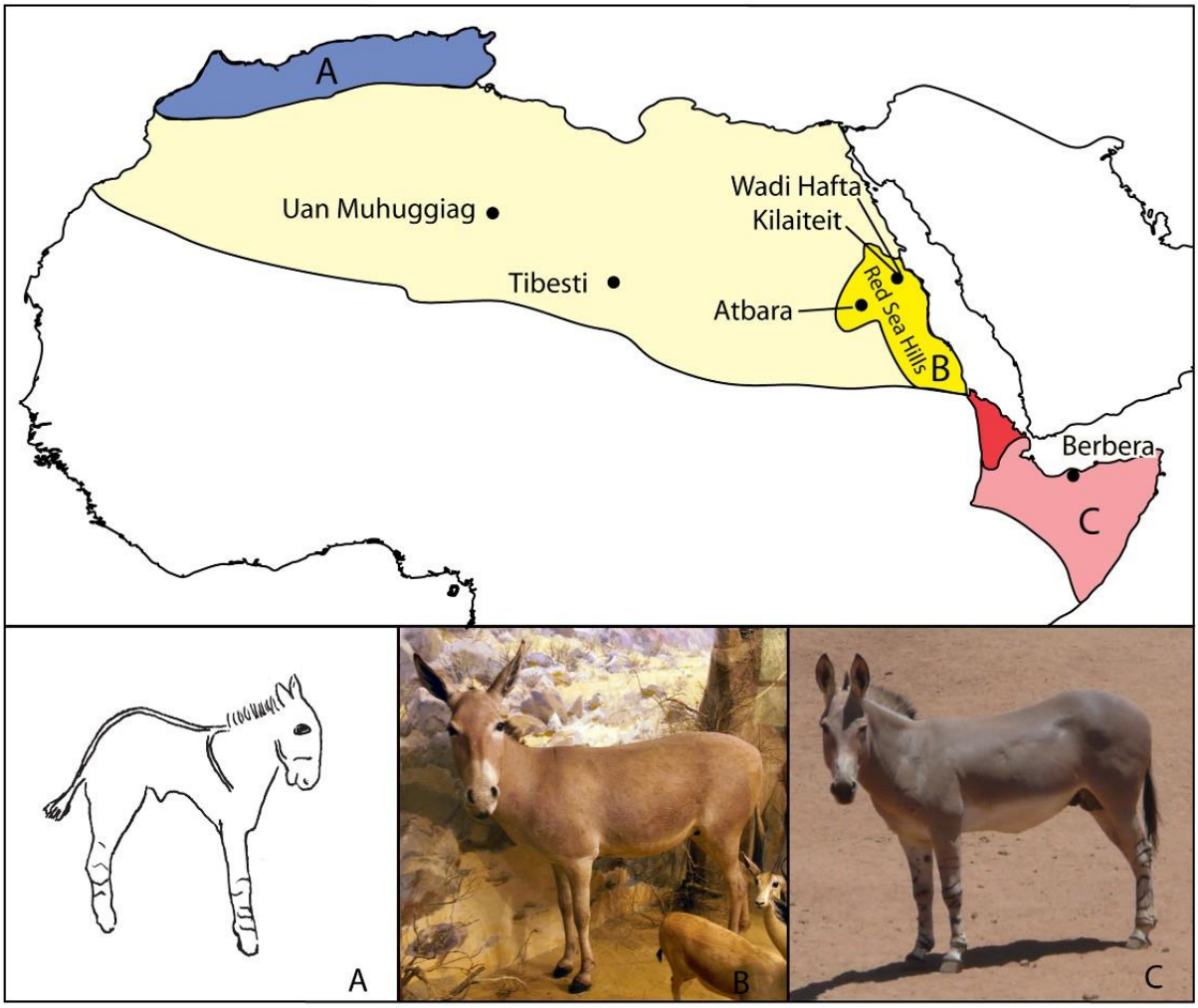


Figure 1 - Map showing the distribution of (a) ancient Atlas wild ass, (b) historic Nubian wild ass and (c) historic Somali wild ass, with drawings or photos of each animal depicted below. The hypothesized extended range of ancient African wild ass across North Africa is indicated in light yellow. The locations of ancient and historic populations are identified within the range of ancient wild ass in blue (Atlas), yellow (Nubian) and pink (Somali). Modern Somali wild ass distribution is shown on the map in red. The locations for all successfully analysed samples are indicated. Image credits: (a) drawn from El Richa image in Muzzolini, 2000, (b) photo with permission of Powell-Cotton Museum, (c) photo Tom Pilgram.

There has also been debate over whether the African wild ass once ranged into western Asia. This is complicated by a lack of reliable historic documentation of *E. africanus* in the region and difficulties in morphological discrimination between *E. africanus* and *E. hemionus*. In addition, the recent discovery that early dynastic Egyptian donkeys used for transport at Abydos are morphologically indistinguishable from the African wild ass (Rossel *et al.* 2008) raises the possibility that faunal specimens attributed to wild ass in Asia could be derived from early domestic donkeys.

Historic populations of wild ass have been distinguished phenotypically by size, and the presence or absence of distinctive leg stripes and shoulder crosses (Groves 1986) (Fig. 1). Osteological differences in size and cranial morphology have also been documented (Groves 1986), but assessment of variability among these wild populations is not straightforward because there are only a few skeletons available in museum collections. Additional samples are virtually impossible to procure because Somali wild asses are critically endangered today, with perhaps as few as 600 individuals left in the wild (Moehlman 2002), and Nubian wild asses have only been infrequently sighted since the 1970s and are therefore considered possibly extinct (Moehlman 1992).

The small size and remote distribution of remaining populations also explains why little is known about the genetics of contemporary African wild ass. Prior to this study, there were only five published sequences in Gen-Bank, three from Somali wild ass and two from individuals tentatively identified as Nubian wild ass (Beja-Pereira *et al.* 2004).

To better understand variability in *E. africanus* populations across their former African ranges, additional samples of the remaining Somali wild ass populations as well as ancient DNA (aDNA) data on historic and ancient African wild ass have been obtained. As will be shown, archaeological and museum specimens represent an invaluable genetic repository for African wild ass.

In light of the genetic research that indicates an African origin for both clades of domestic donkey (Beja-Pereira *et al.* 2004; Marshall 2007), the goal of our study was to investigate African settings for domestication of the donkey and to test the current hypotheses that (i) Nubian wild asses were ancestors of Clade 1 domestic donkeys and (ii) a relative of the Somali wild ass was the ancestor of donkeys of Clade 2 (Beja-Pereira *et al.* 2004). We used aDNA methods to analyze 12 ancient samples from archaeological sites in northeast Africa and Yemen, ranging in age from 3000 years ago to the early Holocene, in addition to nine tissue samples from all known historic Nubian skeletons and two Somali wild ass museum specimens collected between 1880 and 1950 (supplementary material, Table 7). We also collected and analyzed 33 fecal and skin samples

from Somali wild ass populations in Ethiopia and Eritrea. These mitochondrial DNA (mtDNA) data from modern, historical and ancient specimens were combined with previously published sequences for network and phylogenetic analysis.

Material and Methods

Holocene archaeological (n = 12) and historic museum (n = 11) samples of Nubian wild ass, Somali wild ass and donkey were obtained for this study (supplementary material, Table 7) through the comprehensive evaluation of a significant portion of all specimens in existence including: annual camel-based surveys of critically endangered African wild ass conducted by the International Union for Conservation of Nature (IUCN), approximately 12 skeletons of African wild ass held in world museums, and a survey of isolated donkey bones from African archaeological sites. Appropriate permits were obtained for all specimens including CITES permits for all wild ass specimens owing to their status as critically endangered (see electronic supplementary material, Background for more historical and taxonomic information on the historic samples).

Fecal samples from Somali wild ass from Ethiopia (n = 6) and Eritrea (n = 27) were collected across the species habitat range after observation of the animal (supplementary material, Table 7). Each sample was stored in white paper envelopes, dried for 24 h, and shipped to CIBIO-Universidade do Porto. In addition, dried skin from five skeletons of animals that died during the drought of 2006 in Eritrea was used for DNA extraction.

Samples were analyzed using standard precautions for working with ancient DNA. At the University of Florida and Harvard University, analyses were performed in laboratories dedicated to ancient DNA work in which no previous work on equids had been performed. Ground bone samples were extracted with two different methods, one based on DNA binding to silica (Rohland & Hofreiter 2007) and one using phenol/chloroform extraction (Kemp *et al.* 2007). Only one sample and the accompanying extraction blank were processed at a given time. Samples that yielded DNA were re-extracted with a minimum of one other sample being processed in between the first and second extraction of the positive samples. Museum tissue, dry skin and fecal samples were extracted with the Qiagen DNeasy tissue kit at the University of Florida and CIBIO.

For the archaeological samples, primers were designed to amplify segments of 56–158 base pairs (bps) of the most variable regions of the control region that specifically distinguish between domestic donkey clades (supplementary material, Table 8). Museum tissue, dry skin and fecal samples were amplified in three to four overlapping segments ranging from 158 to 308 bps in

length. PCR conditions were as follows: 25 ml reaction with 1x manufacturer's PCR buffer, 2.5 mM Mgþþ, 200 mM each dNTP, 1 mM of each primer, 1.5 mg BSA and 1 unit Amplitaq Gold DNA polymerase or a 25 ml reaction with 1x Bioline Short mix and 1 mM of each primer (see electronic supplementary material, table 8 for annealing temperatures).

A minimum of three independent PCR amplifications were performed with each primer pair. PCR amplification products from the ancient archaeological samples were cloned into a TOPO TA vector (Invitrogen) following the manufacturers' recommendations. Eight to 12 colonies from each amplification product were sequenced and analyzed on a Beckman CEQ 8000, following the manufacturers' recommended protocol for sequencing.

Amplification products from the tissue, dry skin and fecal samples were sequenced directly using the forward and reverse primers that were used for the PCR amplification.

Products from a minimum of three independent PCR reactions were sequenced in both directions for the historic samples. Additional details on extraction, amplification and sequencing are available in supplementary material, material and methods. Newly reported equid sequences were used to create both a median-joining network and a phylogeny. These sequences were aligned with Clustal W from MEGA 4 software (Kumar *et al.* 2008) and compared with previously published sequences; E. asinus NC_001788 (Xu *et al.* 1996), DQ44 878-DQ449 023 (Chen *et al.* 2006) and AY569 462-AY569 547 (Beja-Pereira *et al.* 2004). Sequences of 440 bps were used in the network and phylogenetic construction. In cases where ancient or historic sequences were shorter than 440 bps, but identical to previously published sequences, those sample labels are listed along with the identical (full length) previously published sequences on the network and phylogeny. NHML 1939 yielded a sequence of only 204 bps and is a new sequence, so this sample does not appear in either the network or phylogeny. Median-joining networks (Bandelt 1999) were constructed with NETWORK v. 4.5 (http://www.fluxus-engineering.com/). Reticulations were resolved through a maximum-parsimony criterion (Polzin & Daneshmand 2003). Information on the phylogenetic analysis and estimation of time to most recent common ancestor for each clade can be found in supplementary material, material and methods.

Determination of wild or domestic status of the ancient Uan Mughuggiag magnum, Os Carpale III (specimen Verona 3870, articulated with aDNA no.7), was made on the basis of morphometrical analysis. Size-based identifications of domestic versus wild ass are not reliable for the earliest periods of domestication prior to size decrease (Rossel *et al.* 2008), but smaller donkeys can readily be distinguished from wild ass in Africa after ca 4000 cal year BP. Greatest breadth and maximum length measurements were made to the nearest millimeter using a measuring board

and calipers and following conventions established by von den Driesch (von den Driesch 1976) (supplementary material, Table 9). Uan Muhuggiag measurements were compared with those from seven ancient donkeys, nine modern donkeys and 14 wild ass, including three juveniles. The Uan Muhuggiag mandible (Verona 3988, aDNA no.8) was aged using incisor dental eruption and wear following the sequence documented for donkeys (Misk & Semieka 1997).

Results

The collection of samples successfully analyzed in this study covers the presumed range of Nubian and Somali wild ass over north eastern Africa and includes modern, historic and ancient specimens spanning a time depth of 3000 years. In total, three of the 12 ancient samples, 10 of the 11 historic samples and 33 modern Somali wild ass samples were successfully amplified and sequenced for the mitochondrial control region (supplementary material, Tables 7 and 10; all sequences are available through GenBank, HM622626-HM622669). Final analyzed sequences ranged in size from 201 to 440 bps and were composed of amplicons ranging in size from 33 to 440 bps, i.e. some specimens required multiple fragments to construct an informative DNA sequence. Four historic samples were independently processed and their sequences confirmed at the CIBIO Universidade do Porto, and one ancient sample was independently extracted and amplified at Harvard University (supplementary material, Table 7).

Of the ancient samples, one had the maternal genetic signature of horse and is not analyzed further here. This specimen was represented by a single tooth and had been provisionally identified as donkey. In actuality, the specimen may belong to a mule, i.e. offspring of a female horse and male donkey. Two ancient specimens from the Uan Muhuggiag rock shelter in the central Sahara (Di Lernia & Manzi ; Barich 1987; Gautier 1987), a mandible with one permanent incisor erupted and a trapezoid, were also successfully sequenced. An unciform that was articulated with the trapezoid was directly AMS dated at the University of Oxford to 3160–2975 cal BP (supplementary material, Table 11). The Uan Muhuggiag sequences matched an historic Nubian wild ass sequence reported below (NHML1904, electronic supplementary material, Table 10) and fell in Clade 1, which supports an ancestral role for Nubian wild ass within Clade 1. Furthermore, the Uan Muhaggiag sequences were identical to each other, suggesting that both specimens came from a single individual or from maternally related animals.

Morphometric analysis of the Uan Muhuggiag specimens documented that they were more consistent with those of a small domestic donkey than those expected for either adult or juvenile

wild ass, suggesting that the Uan Muhuggiag animal(s) was domestic (supplementary material, Fig. S1).

Nine of the historic samples were from animals identified as Nubian wild ass on phenotypic and geographical grounds. Eight of those specimens yielded five different sequences that fell within Clade 1 (Fig. 2 and supplementary material, Fig. S2). One of the five sequences was new (BSZM 1952). Four samples, two pairs of mother and fetus, had identical sequences that matched a haplotype also found in domestic donkeys.

They were collected from two areas in close proximity in the Red Sea Hills. Another match to a domestic donkey haplotype was found in an animal from the Tibesti Mountains of the Sahara (RMCA31155). Significantly, the sequence from one sample from the Atbara region in Sudan (NHML1904) exactly matched a modern sequence from the eastern Sudan that had been tentatively identified as Nubian wild ass [H6; Beja-Pereira *et al.* (2004)] suggesting that Nubian wild ass maternal lineages survived at least until the last decade in the eastern Sudan.

The geographical breadth of the successfully assayed specimens confirms the ancient range of the Nubian wild ass in Sudan and northern Eritrea and the presence of animals of Clade 1 not only east, but also west, of the Nile River as far as the central Sahara. The ninth historic sample (BSZM1963) attributed to Nubian wild ass on phenotypic grounds had a sequence identical to a haplotype found in domestic donkeys of Clade 2 (Fig. 2 and supplementary material, Fig. S2).

The 33 modern Somali wild ass specimens fell in the same clade as previous Somali wild ass specimens (WH1-4). This clade is well separated from the domestic donkey Clades 1 and 2 and is clearly not ancestral to either clade (Fig. 2 and supplementary material, Fig. S2). Only four new haplotypes were found in the 33 specimens analyzed and haplotype diversity of the Somali wild ass clade is only 0.7417+0.0444 (compared with 0.9309+0.0102 and 0.8212+0.0268 for Clades 1 and 2, respectively), suggesting that the genetic variability in present-day Somali wild ass is low. The new haplotypes are found in both Eritrea and Ethiopia, and show no geographical structure. The single historic Somali specimen that was successfully amplified came from Berbera, Somalia. Collected around 1886, it showed a sequence identical to that of one of the new Somali wild ass haplogroups (WH1) from Eritrea and Ethiopia (Fig. 2 and supplementary material, Fig. S2 and Table S1). This result demonstrates a degree of historical continuity in the mitochondrial variability of Somali wild ass within the region over the last 120 years. We also calculated the coalescence time of each clade, i.e. the time to the most recent common ancestor (TMRCA), as follows: Clade 1: 406 000 years ago (95% confidence interval 105 400–811 300 years), Clade 2: 334 600 years ago (95% confidence interval 86 100–661 300 years), Somali wild

ass clade: 359 500 years ago (95% confidence interval 57 600–770 800 years). Although there may be some uncertainty in the dates owing to time dependency (Ho & Larson 2006), these dates clearly predate the domestication time for donkey of approximately 5000 years ago.

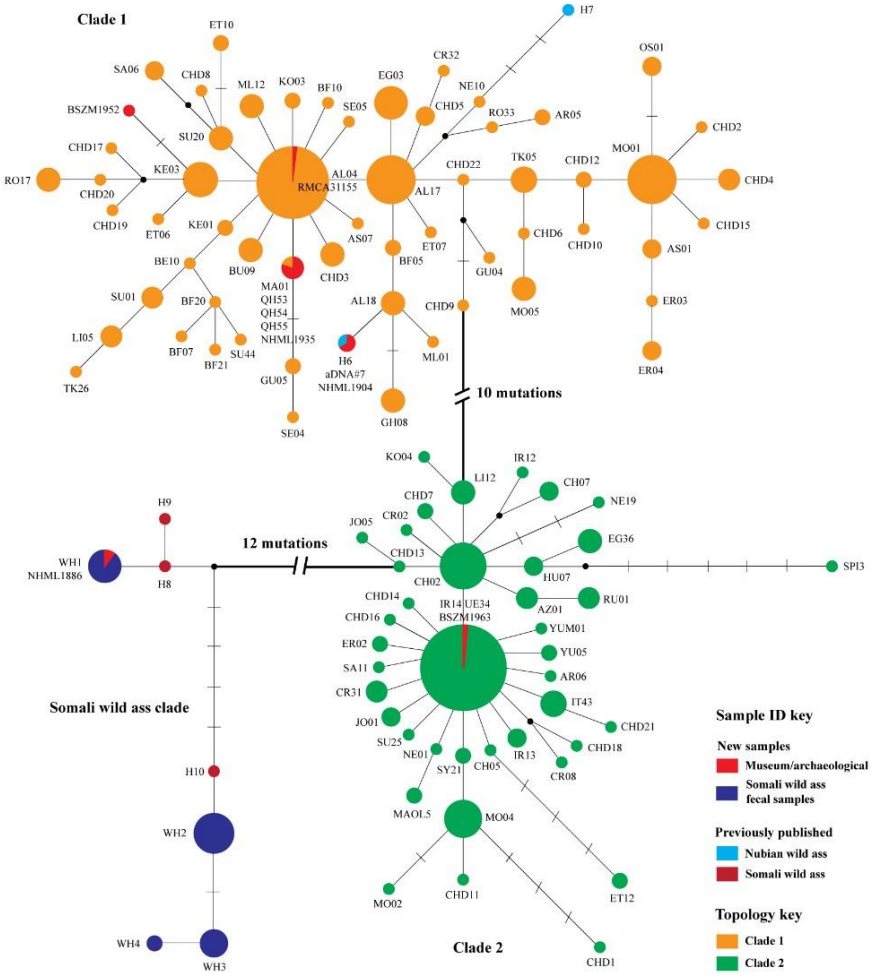


Figure 2 - Reduced median-joining network of 108 mtDNA haplotype sequences from domestic and wild asses. Colored circles represent sampled haplotypes, while black dots represent hypothesized, unsampled haplotypes. Size of the circle is proportional to haplotype frequency and branch length is proportional to number of mutations.

Discussion

Ancestors for the donkeys

The diversity and geographical variability in historic and ancient DNA together with information on modern donkey mitochondrial genetic variation provide new insights into relations among ancestral wild ass, relations of wild ass to domestic haplogroups, and the process of donkey

domestication. Our results demonstrate that Nubian and Somali wild asses are mitochondrially distinct. Furthermore, we show that the historic Nubian wild asses and domestic donkeys of Clade 1 are almost indistinguishable on the basis of mtDNA; five of our historic Nubian wild ass samples had haplotypes identical to domestic donkeys of Clade 1. It is probable that the identical wild ass sequences represent survival of the originally domesticated maternal haplotypes in the wild population, although we cannot rule out the possibility that they were introduced into wild herds by feral female donkeys. Historic specimens collected by naturalists over the last two centuries verify a northern Sudanese and Eritrean distribution of Nubian wild ass in northeast Africa, but the Uan Muhuggiag and Tibesti data suggest that donkeys of Clade 1 and/or Nubian wild asses were present as far west as the central Sahara in late (pre)historic times.

When our results are combined with 98 previously published haplotypes, the topology of the network (Fig. 2) provides some interesting perspectives on domestication processes. The Clade 1 topology resembles that found in European and Asian domestic pigs (Larson *et al.* 2007) and domestic and wild reindeer (Røed *et al.* 2008), with several smaller nodes and wild animals interspersed with domestic. It presents an interesting example of survival of wild populations during the process of domestication giving rise to indistinct differences at the mtDNA level between wild and domestic individuals. This process has been proposed for horse, dog, pig and reindeer (Jansen *et al.* 2002; Savolainen *et al.* 2002; Larson *et al.* 2005; Røed *et al.* 2008), but we are able to verify this conclusion for donkeys since we have multiple cases of mtDNA sequences shared by wild and domestic specimens. The Clade 1 topology is consistent with a scenario whereby the Nubian wild ass was domesticated in several areas and/or over an extended long period, with multiple recruitments from the wild, similar to the domestication process suggested for dogs and goats (Vilà *et al.* 2001; Naderi *et al.* 2008). The much broader distribution of Nubian wild ass in former times and likely domestication by cattle herders who ranged widely over the Sahara from as early as 8900–8400 cal BP provide geographical, social and temporal contexts for these processes (Hanotte *et al.* 2002; Marshall & Hildebrand 2002; Gifford-Gonazalez 2005). Moreover, the use of 'morphologically wild' donkeys for heavy transport at Abydos in ancient Egypt ca 3000 BC (Rossel *et al.* 2008), has already illustrated a slow process of donkey domestication with late morphological and genetic change.

Introgression with wild ancestors is especially probable among donkeys because they are not herd animals and are not intensively managed by African pastoralists (Marshall 2007). Pastoralists who keep donkeys for transport particularly value strength and reproductive potential in their animals and recruit both males and female donkeys to their herds (Marshall 2007).

Furthermore, capture of wild ass through trapping is historically documented and illustrated in African rock art (Baker 1867; Murray 1935; Nicolaisen 1963), but is indiscriminate with respect to sex (for additional information, see supplementary material, Background).

We conclude that donkeys of Clade 1 have a long history in the Sahara, that a Nubian wild ass was their ancestor, and that it is probable there was interbreeding between wild and domestic forms over a long period of time with recruitment of several maternal haplotypes from the wild.

The contribution of the Somali wild ass to the domestic gene pool

Our extended mitochondrial dataset from free-living Somali wild ass shows that Somali wild ass are distinct from Nubian wild ass and domestic donkeys of both Clades 1 and 2 (Fig. 2 and supplementary material, Fig. S2). Given the extensive haplotype networks found in Clades 1 and 2, it is surprising to find so few Somali wild ass haplotypes after increasing the sample size by an order of magnitude. The low variation and large sample size of Somali wild ass make it unlikely that additional lineages will be identified and, thus, make Somali wild ass a less probable candidate for the ancestor of Clade 2 than previously thought (Beja-Pereira *et al.* 2004; Chen *et al.* 2006; Vilà *et al.* 2006). Furthermore, the equal distance of the three major clades to each other diminishes the possibility that the ancestor of Clade 2 lies in either of the other two clades. The very old coalescence times of the three clades reflect the long period of time before donkey domestication and suggest that substantial genetic structuring, fragmentation and/or geographical isolation of wild ass mitochondrial variation may have developed prior to domestication. As a result we cannot rule out the possibility that wild ass in northeast Africa may have had additional, yet unrecognized, genetic substructure and particularly that Clades 1 and 2 may both have Nubian-like wild ass ancestors (see supplementary material, Background for additional details).

In addition, the observation that wild ass/donkey mitochondrial variation may have undergone significant reductions over time also raises the possibility that the ancestor of Clade 2 belonged to an extinct population.

Archaeological data suggest the Holocene ranges of African wild ass were substantially more extensive, the presence of wild ass in the central and eastern Sahara being evidenced by rock engravings (Lutz & Lutz 1995) and skeletal remains (Gautier 1987; Peters & Pollath 2004; Marshall 2007), which is consistent with our genetic results. Thus, there are several possibilities for the geographical origin of the wild ancestor of Clade 2. In addition to northeastern Africa, candidates include the ancient range of the Atlas wild ass in the Maghreb (Groves 1986) and the

coast of Yemen, where specimens identified as early domestic donkeys or wild ass have been excavated (Cattani & Bokonyi 2002; Marshall 2007); however, our ancient samples from these regions did not yield genetic material. Additional aDNA research in Africa and Asia as well as Y chromosome or nuclear genetic data on donkeys and extant African wild ass are needed to pinpoint the locus of domestication of Clade 2.

Patterns of domestication and conservation implications

The findings presented in this study clarify the role of the Nubian wild ass in the domestication of the donkey but raise new questions regarding the second ancestor for the donkey. Evidence for domestication of several Nubian haplotypes, multiple recruitments from the wild, and ongoing gene flow in Clade 1, contrasts with a simpler domestication process starting from fewer ancient founders for Clade 2. These distinct patterns fit with recent research on other livestock species showing multiple domestication events with differing histories, social contexts and timelines (Hanotte *et al.* 2002; Beja-Pereira *et al.* 2004; Bradley & Magee 2006; Vilà *et al.* 2006; Zeder *et al.* 2006; Larson *et al.* 2007; Naderi *et al.* 2008; Røed *et al.* 2008).

Our findings also have several implications for conservation. (i) Nubian wild asses are distinct from Somali wild asses based on mtDNA, a result that indicates the need for separate management of Nubian and Somali populations. (ii) The finding that maternal lineages of the Nubian wild ass may have survived in the eastern Sudan until the 1990s implies that Nubian wild asses are not extinct or became extinct very recently, and reinforces the need for surveys and management plans for eastern Sudan and northern Eritrea. (iii) Extant Somali wild ass in Eritrea and Ethiopia shows an absence of geographical structuring of genetic variation as well as low haplotype diversity, which may reflect past bottlenecks in ancestral populations in which short-term crashes wiped out many haplotypes. These results suggest that captive breeding populations of Somali wild ass already sample much of the available mitochondrial diversity. Our findings represent a valuable contribution to debates regarding variability, phylogeny and management of extant but critically endangered African wild ass (Groves 1986; Yalden *et al.* 1986; Moehlman 2002). Our research also underscores the need for further studies of the nuclear and Y chromosomal DNA of extant populations and for more specimens for aDNA analysis.

Acknowledgments

This research was supported by National Science Foundation grant BCS-0447369 (FM), FCT grants (SFRH/BPD/26802/2006, SFRH/BPD/17822/2004, and PTDC/BIA/BDE/64111/2006 (ABP) and grants from Wildlife Trust, Saint Louis Zoo, Basel Zoo, Liberec Zoo and Sea World and Busch Gardens Conservation Fund (PDM). The authors would like to thank four anonymous reviewers for many useful comments. We thank V. Costa for help during fecal DNA extraction and T. Schilling for assistance with figure 1. We are indebted to F. Alhaique the late A. Bietti, L. Chaix, L. Khalidi, C. Lewis, R. Meadow, W. van Neer, T. Webber, W. Wendelen and to the institutions and individuals listed in supplementary material, table S1 who generously provided material or assisted in collection of samples for this study.

References

Baker S. (1867) *The Nile tributaries of Abyssinia*. Macmillan, London, UK.

Bandelt H.-J., Forster, P., Rohl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37-48.

Barich B. (1987) Archeology and environment in the Libyan Sahara, 1978-1983 In: *British Archeological Reports* (Oxford, UK.

Beja-Pereira A., England P.R., Ferrand N., Jordan S., Bakhiet A.O., Abdalla M.A., Mashkour M., Jordana J., Taberlet P. & Luikart G. (2004) African origins of the domestic donkey. *Science* **304**, 1781.

Bradley D.G. & Magee D.A. (2006) Genetics and origins of domestic cattle. In: *Documenting domestication* (eds. by Zeder M, Bradley D, Emshwiller E & Smith BD), pp. 317-28. University of California Press, Berkeley, California, USA.

Cattani M. & Bokonyi S. (2002) Ash-Shumah: an early Holocene settlement of desert hunters and mangrove foragers in the Yemeni Tihamah In: *Essay on the late Prehistory of the Arabian Peninsula* (eds. by Cleuziou S, Tosi M & Zarins J), pp. 31-53. Instituto Italiano per l'Afirca e l'Oriente, Rome.

Chen S.Y., Zhou F., Xiao H., Sha T., Wu S.F. & Zhang Y.P. (2006) Mitochondrial DNA diversity and population structure of four Chinese donkey breeds. *Animal Genetics* **37**, 427-9.

Clutton-Brock J. (1992) *Horse power: a history of the horse and the donkey in human societies*. Harvard University Press.

Di Lernia S. & Manzi G. *Sand, stones and bones*. All'Insegna del Giglio, Florence, Italy.

Gautier A. (1987) The archeological sequence of the Acacus. In: *Archeology and environment in the Libyan Sahara, 1783-1983* (ed. by Barich BE), pp. 283-308. British Archeological Reports, Oxford, UK.

Gifford-Gonazalez D.P. (2005) *Pastoralism and its consequences*. Blackwell, Cambridge, UK.

Groves C.P. (1986) The taxonomy, distribution, and adaptations of recent equids. In: *Equids in the ancient world* (eds. by Meadow RH & Uerpmann H-P), pp. 11-65, Wiesbaden.

Hanotte O., Bradley D.G., Ochieng J.W., Verjee Y., Hill E.W. & Rege J.E. (2002) African pastoralism: genetic imprints of origins and migrations. *Science* **296**, 336-9.

Hassan F. (1993) Town village in ancient Egypt. In: *The archeology of Africa: food, metals and towns* (pp. 551-69. Routledge, London, UK.

Ho S.Y.W. & Larson G. (2006) Molecular clocks: when timesare a-changin'. *Trends in Genetics* **22**, 79-83.

Jansen T., Forster P., Levine M.A., Oelke H., Hurles M., Renfrew C., Weber J. & Olek K. (2002) Mitochondrial DNA and the origins of the domestic horse. *Proceedings of the National Academy of Sciences* **99**, 10905-10.

Kemp B.M., Malhi R.S., McDonough J., Bolnick D.A., Eshleman J.A., Rickards O., Martinez-Labarga C., Johnson J.R., Lorenz J.G., Dixon E.J., Fifield T.E., Heaton T.H., Worl R. & Smith D.G. (2007) Genetic analysis of early holocene skeletal remains from Alaska and its implications for the settlement of the Americas. *Am J Phys Anthropol* **132**, 605-21.

Kumar S., Nei M., Dudley J. & Tamura K. (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* **9**, 299-306.

Larson G., Albarella U., Dobney K., Rowley-Conwy P., Schibler J., Tresset A., Vigne J.-D., Edwards C.J., Schlumbaum A., Dinu A., Bălăçsescu A., Dolman G., Tagliacozzo A., Manaseryan N., Miracle P., Van Wijngaarden-Bakker L., Masseti M., Bradley D.G. & Cooper A. (2007) Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Sciences* **104**, 15276-81.

Larson G., Dobney K., Albarella U., Fang M., Matisoo-Smith E., Robins J., Lowden S., Finlayson H., Brand T., Willerslev E., Rowley-Conwy P., Andersson L. & Cooper A. (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* **307**, 1618-21.

Lutz R. & Lutz G. (1995) *Secret of the desert*. Golf, Innsbruck, Austria.

Marshall F. (2007) *Rethinking Agriculture: Archeological and Ethnoarcheological perspectives*. Left Coast Press, Inc., Walnut Creek, California, USA.

Marshall F. & Hildebrand E. (2002) Cattle Before Crops: The Beginnings of Food Production in Africa. *Journal of World Prehistory* **16**, 99-143.

Misk N.A. & Semieka M.M.A. (1997) In: *Equine Practice*, pp. 23-9.

Moehlman P.D. (1992) African wild asses (Equus africanus Fitzinger). In: *Zebras, asses and horses. An action plan for the conservation of wild equids* (ed. by Duncan P), pp. 15-7. IUCN, Gland, Switzerland.

Moehlman P.D. (2002) *Equids: Zebras, Asses and Horses: Status Survey and Conservation Action Plan.* IUCN/SSC Equid Specialist Group Gland, Switzerland and Cambridge, United Kingdom.

Murray G.W. (1935) *Sons of Ishmael*. Routledge, London, UK.

Naderi S., Rezaei H.-R., Pompanon F., Blum M.G.B., Negrini R., Naghash H.-R., Balkız Ö., Mashkour M., Gaggiotti O.E., Ajmone-Marsan P., Kence A., Vigne J.-D. & Taberlet P. (2008) The goat domestication process inferred from large-scale mitochondrial DNA

analysis of wild and domestic individuals. *Proceedings of the National Academy of Sciences* **105**, 17659-64.

Nicolaisen J. (1963) *Ecology and culture of the Pastoral Tuareg*. The National Musum of Copenhagen, Copenhagen, Denmark.

Peters J. & Pollath N. (2004) Holocene faunas of the eastern Sahara: zoogeographical and palaeoecological aspects. In: *The future from the past* (eds. by Lauwerier RCGM & Plug I), pp. 34-50. Oxbow Books, Oxford, UK.

Polzin T. & Daneshmand S.V. (2003) On Steiner trees and minimum spanning trees in hypergraphs. *Operations Research Letters* **31**, 12-20.

Røed K.H., Flagstad Ø., Nieminen M., Holand Ø., Dwyer M.J., Røv N. & Vilà C. (2008) Genetic analyses reveal independent domestication origins of Eurasian reindeer. *Proceedings of the Royal Society B: Biological Sciences* **275**, 1849-55.

Rohland N. & Hofreiter M. (2007) Comparison and optimization of ancient DNA extraction. *Biotechniques* **42**, 343-52.

Rossel S., Marshall F., Peters J., Pilgram T., Adams M.D. & O'Connor D. (2008) Domestication of the donkey: timing, processes, and indicators. *Proc Natl Acad Sci U S A* **105**, 3715-20.

Savolainen P., Zhang Y.P., Luo J., Lundeberg J. & Leitner T. (2002) Genetic evidence for an East Asian origin of domestic dogs. *Science* **298**, 1610-3.

Smith D.G. & Pearson R.A. (2005) A review of the factors affecting the survival of donkeys in semi-arid regions of sub-Saharan Africa. *Trop Anim Health Prod* **37 Suppl 1**, 1-19.

Starkey P. (2000) The history of working animals in Africa. In: *The origins and development of African livestock: archeology, genetics, linguistics and ethnography* (eds. by McDonald KC & Blench RM), pp. 478-502. University College London Press, London, UK.

Uerpmann H.-P. (1991) Equus africanus in Arabia. In: *Equids in the Ancient World* (eds. by Meadow RH & Uerpmann H-P), pp. 12-33. Ludwig Reichert Verlag, Wiesbaden.

Vilà C., Leonard J.A. & Beja-Pereira A. (2006) Genetic documentation of horse and donkey domestication In: *Documenting domestication* (eds. by Zeder MA, Bradley DG, Emshwiller E & Smith BD), pp. 342-53. University of California Press, Berkeley, California, USA.

Vilà C., Leonard J.A., Götherström A., Marklund S., Sandberg K., Lidén K., Wayne R.K. & Ellegren H. (2001) Widespread Origins of Domestic Horse Lineages. *Science* **291**, 474-7.

Vila E. (2002) Data on equids from late fourth and third millennium sites in Northern Syria. In: *Equids in time and space* (ed. by Mashkour M), pp. 101-23. Oxbow Books, Oxford, UK.

von den Driesch A. (1976) *A guide to the measurement of animal bones from archeological sites*. Peabody Museum Bulletins, Harvard University Cambridge, MA, USA.

Xu X., Gullberg A. & Arnason U. (1996) The complete mitochondrial DNA (mtDNA) of the donkey and mtDNA comparisons among four closely related mammalian species-pairs. *Journal of Molecular Evolution* **43**, 438-46.

Yalden D.W., Largen M.J. & Kock D. (1986) Catatalogue of the Mammals of Ethiopia 6. *Monitore Zoologico Italiano. Supplemento* **21**, 31-103.

Zeder M.A., Bradley D.G., Emshwiller E. & Smith B.D. (2006) *Documenting domestication*. University of California Press, Berkeley, California, USA.

Supplementary Material

Materials and Methods

Preparation of samples - At the University of Florida extractions were performed in a laboratory in a building separate from the laboratory where the samples were cloned and sequenced. This extraction laboratory has positive air pressure and had never been used for any equid DNA studies. Disposable gloves, hairnets and lab coats were used when handling samples and reagents were either certified DNA free or tested by PCR to be free of exogenous DNA prior to extraction. All surfaces were cleaned with 20% dilute bleach, i.e. 1.2% sodium hypochlorite, in between samples. One to two mm of the outer layer of bone and any discolorations were removed with a new sterile scalpel blade prior to extraction. Teeth were split and cleaned pieces from the inner, bottom half of the tooth used. The mortar and pestle used to grind samples was cleaned in between samples first with detergent and water, followed by submersion in Nochromix sulfuric acid overnight, then rinsed and soaked in 20% dilute bleach for 15 minutes and finally dried after 3 rinses in DNA free water.

Historic museum tissue and dry skin samples were rinsed in DNA free water and rehydrated in PBS overnight at 42^{o}C. NHML 1886, NHML 1904, NHML 1935, and PCM 55 were processed both at the University of Florida and Universidade do Porto. For the fecal samples, a small portion from the external side of the sample was scraped into a 2 ml screw cap tube and 1 ml of PBS was added to the tube. Samples were left overnight at room temperature and the precipitate was used to extract DNA. Dry skin tissue from animals that died in 2006 were prepared as the fecal samples.

DNA extraction - Bone and teeth samples were extracted using the following two protocols: Method A [slightly modified from Rohland and Hofreiter (2007)]: DNA from 100-125 mg of bone or tooth powder was extracted overnight in 2.5 ml extraction buffer containing 0.45M EDTA pH 8.0 and 0.25 mg proteinase K/ml. DNA was then bound to silica [prepared as in (Boom *et al* 1990)] at pH 4 after adding 10 ml of a buffer containing 5.0M guanidine thiocyanate, 0.05M Tris and 0.25M sodium chloride. After washing the silica twice with a solution of 51.3% ethanol, 125 mM sodium chloride, 10 mM Tris and 1 mM EDTA, the DNA was eluted in 50 µl TE pH 8.0. Method B [slightly modified from Kemp *et al.* (2007)]: 200-300 mg of bone or tooth powder was decalcified in 2 ml 0.5 M EDTA pH 8.0 for 3-6 days. Then 3 mg proteinase K was added and the sample incubated at 55 degrees Celsius for 16-24 hours. After the addition of an equal volume of DNA-free water, the samples were extracted twice with phenol/chloroform/isoamyl alcohol (25/24/1)

and once with chloroform/isoamyl alcohol (24/1). DNA was precipitated overnight with isopropanol in the presence of 0.83 M ammonium acetate, the precipitate washed with 80% ethanol and dissolved in 100 µl TE pH 8.0. Museum tissue, dry skin, and fecal samples were extracted with the Qiagen DNeasy kit following the manufacturer's recommendations with a final elution volume of 75 µl.

PCR amplification - Primers were designed with the Primer3 program (Rozen & Skaletsky 2000). The primers are named based on the donkey sequence NC_001788 (Xu *et al.* 1996) deposited in Genbank. Primer pairs L15462/H15533 and L15492/H15600 overlap to cover one highly variable region and primer pairs L15669/H15791 and L15769/H15844 cover another highly variable region with a slight gap. The primers are situated in a more conserved region. Primer pairs L15669/H15724 and L15722/H15791 amplify shorter fragments and were used for verification. These primers were used for the archaeological samples. PCR conditions were as follows: After an initial 7 minute heating step at 94ºC, the reactions were cycled 50 times at 94º for 30 seconds, at the appropriate annealing temperature for 1 minute, and at 72º for 1 minute, followed by a final extension of 7 minutes. Sample volume was generally 1 µl, but different volumes between 2 µl and 1 µl of a 1:10 dilution were also tested. Nested and semi-nested amplifications were used to obtain enough material for sequencing from sample NHM 1939 and to confirm the sequence of a 122bp segment from sample aDNA#8. For nested and semi-nested amplifications, the number of cycles was reduced to 45 and the times for annealing and extension in each cycle reduced to 45 seconds. Amplification products were visualized and photographed after electrophoresis on 1.5% agarose gels containing 50 µg ethidium bromide/100ml agarose solution.

Sequence analysis - Phylogenetic relationships among haplotype sequences were estimated by Bayesian analysis with MrBayes 3.1.2 (Ronquist & Huelsenbeck 2003) using a horse sequence (H11, AY569548) as an outgroup. The prior best-fitting nucleotide substitution model was HKY + I + G model, which was selected by Modeltest version 3.7 (Posada & Crandall 1998) based on likelihood ratio tests. Two independent analyses starting from different random trees were performed, and four MCMC chains were run for 5 million generations with sampling every 100 generations. Five thousand trees were discarded as burn-in, after checking for convergence. Diversity measures were calculated using Arlequin 3.11 (http://lgb.unige.ch/arlequin/). The times to the most recent common ancestors (TMRCAs) of the three major clades obtained in the

phylogenetic analysis were estimated using Bayesian relaxed phylogenetic approach (uncorrelated lognormal model) implemented in BEAST v1.5.0 (Drummond *et al.* 2006; Drummond & Rambaut 2007). We did not use a published substitution rate, but calculated our own empically-based rate of 3.6 X 10^{-8} per generation. Specifically, the ages of the sequences (tip dates) obtained from museum specimens and the radiocarbon dating of the ancient DNA samples were incorporated as calibration points. To avoid the need for large extrapolation because the tip dates (maximum at ~ 3100 years) are much younger than those of internal nodes, an additional calibration point was provide by setting a normal prior distribution centered at 2.95 million years with a standard deviation of 0.4845 million years for the root of tree (treeModel.rootHeight parameter). This corresponds to a central 95% range of 2-3.9 million years for the divergence time of the horse and the donkey, derived from paleontological and molecular evidence (Vilà *et al* 2001). In sum, the collection dates of the museum specimens (~200 yrs ago to present time), the radiocarbon date of the ancient specimen (~3100 yrs), and the horse-donkey divergence date (2.95 mya) were included as calibration points to estimate the substitution rate in the phylogeny used in our clade date estimation. The nucleotide substitution model HKY + I + G (as in the phylogenetic analysis) was used in MCMC analysis. Parameters were sampled at every 100 generations over a total of 10 million generations after discarding 1 million burn-in generations. The analysis was repeated and the samples from two runs were combined. Convergence of the sampled parameters was checked using Tracer v1.4.1 (Rambaut 2007), which showed the effective sample sizes of parameters were, at least, > 300.

Additional historical and taxonomic information on the historic wild ass samples that successfully amplified

All but one of the animals studied were collected from the wild between the 1880's and 1930's by naturalists, explorers and hunters. Taxonomic identifications were based on phenotypic characteristics and influenced by the locales in which animals were found and taxonomic debates of the day**.** Animals studied all showed phenotypic characteristics of wild ass rather than those of feral donkeys.

Somali wild ass - We successfully obtained genetic material from one historic specimen NHML 1886, the type specimen for Somali wild ass. The distinctive southern population of "Somaliland" or northern Somalia near Berbera was documented in the 1860's by von Heuglin (Groves &

Smeenk 2007). Strongly marked leg stripes, lack of shoulder cross and large size were characteristic of the wild ass of this region (Groves 1986). According to the records of the Natural History Museum, London, NHML 1886 was collected in Somaliland near Berbera in 1886. It is a young animal and probably the individual collected by Menges on the Hekebo Plateau in the 1880's (Groves 1986; Groves & Smeenk 2007)

Nubian wild ass - We successfully obtained genetic data from two historic Nubian wild ass samples from the Atbara, four from Red Sea Hills populations, two from Eritrea and thought to be from the Red Sea Hills, and one from the Sahara.

Atbara population: Descriptions of northern populations of African wild ass in eastern Sudan (Northern Kassala) on the Atbara and nearby tributaries of the Nile go back to the observations of von Heuglin in the 1860's (Groves & Smeenk 2007). These animals were distinguished by the presence of long thin shoulder crosses, lacked marked leg and were smaller than Somali wild ass (Groves 1986). The two specimens that we studied from the Natural History Museum, London were shot by hunter/naturalists on the Atbara River in northern Sudan. NHML 1904 was shot at Nakheila on the Atbara River in 1904 and was presented to the Natural History Museum by N.C. Rothschild. This animal is depicted by Lydekker (Lyddeker 1904, 1916) with a faint shoulder cross and no leg stripes, and its skin and cranium was studied by Groves (1986). NHML 1939 was killed at Nakheila on the Atbara River in the Sudan in 1939 and presented to the Museum by N.C. Rothschild, ex. Inng collection.

Red Sea Hills population: The taxonomic history of the Red Sea Hills population of the eastern Sudan is more complex. This population was documented in the late 1800's and is phenotypically distinctive, characterized by the presence of traces of stripes on the pasterns and fetlocks. Dollman (1935) considered it intermediate between Nubian *africanus* and s*omaliensis* and gave it a separate sub-species designation *Asinus asinus dianae*. *A. a. dianae* is now considered a synonym of *E. a. africanus* from the Atbara (Groves & Smeenk 2007).

Four Nubian wild ass, Powell-Cotton Museum specimens 53, 54, 55 and NHML 1935 from the Natural History Museum in London, were shot and documented in detail by Major Powell-Cotton and his daughter Diana in 1935 on a hunting and collecting trip in remote areas of the Sudanese Red Sea Hills near Wadi Hafta 17°35' N 37° 35'S and Kilaiteit 17°35'N 37° 35'S. PCM 54 is the co-type of *A. a. dianae*. Phenotypically these animals are typical of Red Sea Hill Nubian wild ass

and the mounted specimen and skins at the Powell-Cotton Museum show a faint shoulder cross and traces of stripes on the pasterns and fetlocks (Groves 1986; pers. obs.).

We also studied two phenotypically Nubian specimens thought to originate in the Red Sea Hills, adult female BSZM 1952 and male BSZM 1963. The latter is also known as "Max" (studbook no 1.) and was the breeding stallion of the Nubian herd at Tierpark Hellabrun, Munich. These animals were collected as adults in Eritrea in the 1930's. The records and collection details on these animals were lost, however, during the Second World War. As a result we do not know exactly where in Eritrea animals were collected or whether they came from the same locale. Groves (1986) notes that based on skins, photographs and observations, "Max" and the three adult females with him had the "smudges" or "broken stripes" on the fetlocks characteristic of the Red Sea Hills Nubian population. Nevertheless, a card (apparently not dating to the 1930's) in the Bavarian State Museum Zoological collections associated with BSZM 1963 skeleton "Max" reads Danakil. This is most unlikely to be correct however, because Max lacked the marked leg stripes characteristic of all Somali wild ass. Scholars who have studied these collections have found the specimens to group with wild ass of the Red Sea Hills rather than with the Somali wild ass.

In his cranial analysis of Red Sea Hills specimens, Groves (1986) notes that this population is variable but more similar to the Atbara than Danakil or other Somali populations. Metacarpal metrics show that BSZM 1963 "Max" was a large animal consistent with wild ass rather than recent domestic donkey of the Horn. This animal lies on the tall end of the range for Nubian wild ass, but is smaller than Somali wild ass (Rossel *et al.* 2008). Divé and Eisenmann (1991) research on phalanx morphometrics also clusters BSZM 1963 and BSZM 1952 with wild ass, but notes that some features are similar to Somali. Both postcranial studies demonstrate that 1952 is unusually small, but is characterized by robust morphological features typical of wild ass (Divé & Eisenmann 1991; Rossel *et al.* 2008). Taken together three different morphological studies group BSZM 1963 and 1952 with wild ass, not with feral donkeys. They also suggest that the animals fit with Red Sea Hills Nubian populations, somewhat intermediate in morphometric characters between Atbara Nubian populations and Somali wild ass of the south.

Saharan population: RMCA 31155 in the Museum voor Middenafrika, Tervuren Belgium, was collected from the Tibesti Mountains. Groves (1986) discusses this specimen in his consideration of the existence of Saharan wild ass, indicating that RMCA 31155 is larger than local donkeys but smaller than Nubian wild ass from Sudan and northern Eritrea. He also notes in support of the existence of Saharan wild ass, that the shoulder cross of another Saharan specimen -the animal

captured and tamed by the Tuareg and depicted by Nicolaisen and Nicolaisen (1997) - is of the wild form.

The Atbara, Red Sea Hills and Saharan populations discussed here raise the issue of whether there are distinct sub-species of historic wild ass as some have argued, or whether it is more productive to focus on clinal gradations among populations (Yalden *et al.* 1986). Groves (1986) suggests that a series of distinct populations of wild ass existed historically, amongst which there was periodic gene flow. On the basis of cranial metrics he argues that the Danakil population of Somali wild ass from southern Eritrea is more similar to the southern Somali populations, than is the Red Sea Hills to the Atbara Nubian population. Insufficient data exist to characterize Saharan or Atlas populations.

Discrimination of feral donkeys from wild ass

There is a history of debate regarding incidences of possible confusion of feral donkeys with wild ass. Historic and modern domestic donkeys are significantly smaller than wild ass, though, characterized by rougher, less reflective coats, differences in mane growth (pers. obs.) and long bold shoulder crosses (Groves 1986). Groves (1986) describes the much more marked shoulder cross of the domestic form and details other more subtle distinctions in coloration. In the earlier years of the century *E. taeniopus* was described based on an animal which does not look like any historically known wild ass or donkey and is now considered to be a donkey-wild ass hybrid (Groves & Smeenk 2007).

The question of whether and how often hybridization occurred, and whether hybrids existed in the wild or primarily in domestic donkey herds is a vexed one. The attempts of herders in some regions to breed their female donkeys with male wild ass is well documented (Baker 1867; Murray 1935; Nicolaisen 1963). In general, however African pastoralists do not attempt to control donkey breeding, which contributes to genetic variability among modern domestic donkeys (Marshall & Weissbrod 2009). Whether feral female donkeys survived in wild herds, or female wild ass were ever captured and tamed is less clear. Many scholars think that feral female donkeys are unlikely to have flourished in wild herds (Groves 1986). It is unlikely, but nevertheless possible, that domestic female mitochondrial lineages found their way into wild herds and are represented in the samples studied. Nicolaisen (1963); Nicolaisen and Nicolaisen (1997) do not specify the gender of the Saharan wild ass that he saw captured and tamed by the Tuareg. Most of the Tuareg hunting methods described involve the use of traps and would not discriminate among

sexes, however. The Tuareg also prefer females for carrying water and for riding (Nicolaisen 1963; Nicolaisen & Nicolaisen 1997). Other African pastoralists prefer females for their reproductive potential (Marshall & Weissbrod 2009). It seems likely that in such circumstances wild females would sometimes be added to domestic herds. Breeding among female donkeys and male wild ass was also still encouraged until recently in several regions of Africa (Baker 1867; Murray 1935; Nicolaisen 1963). Such crossbreeding with wild sires has a long tradition in northern Africa being advocated by the Carthaginian author Mago in the 2nd century BC (Richter 1982).

Figures

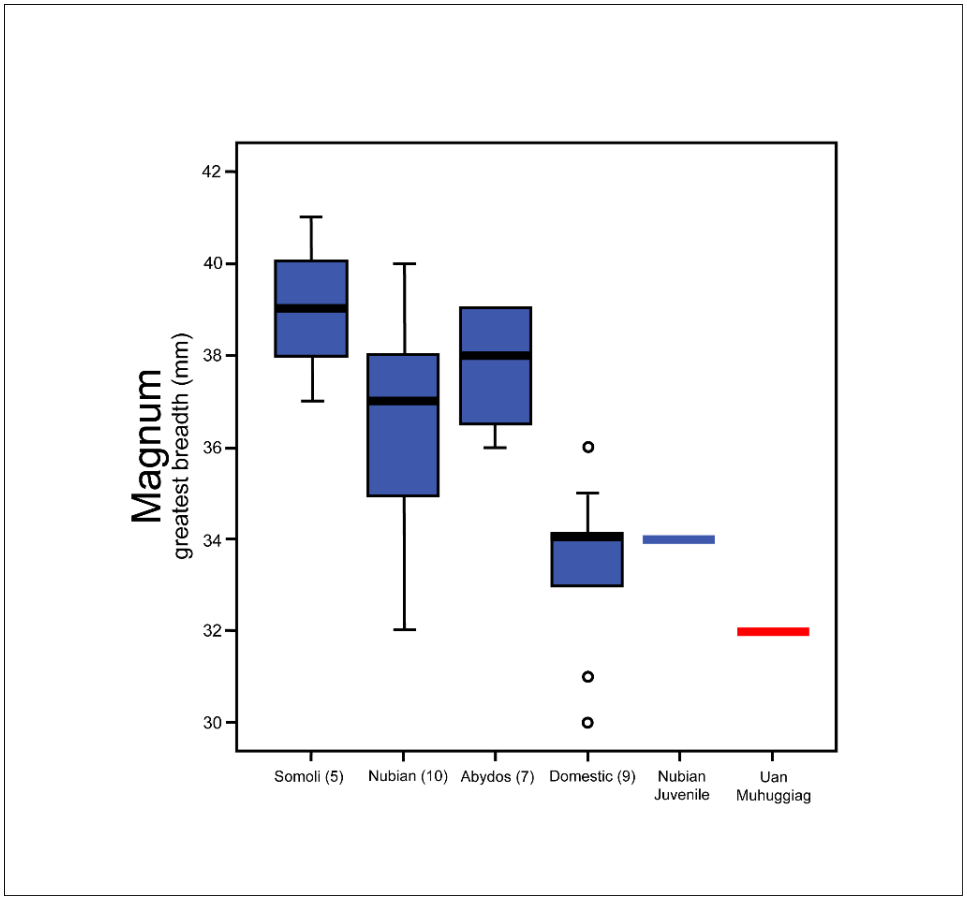
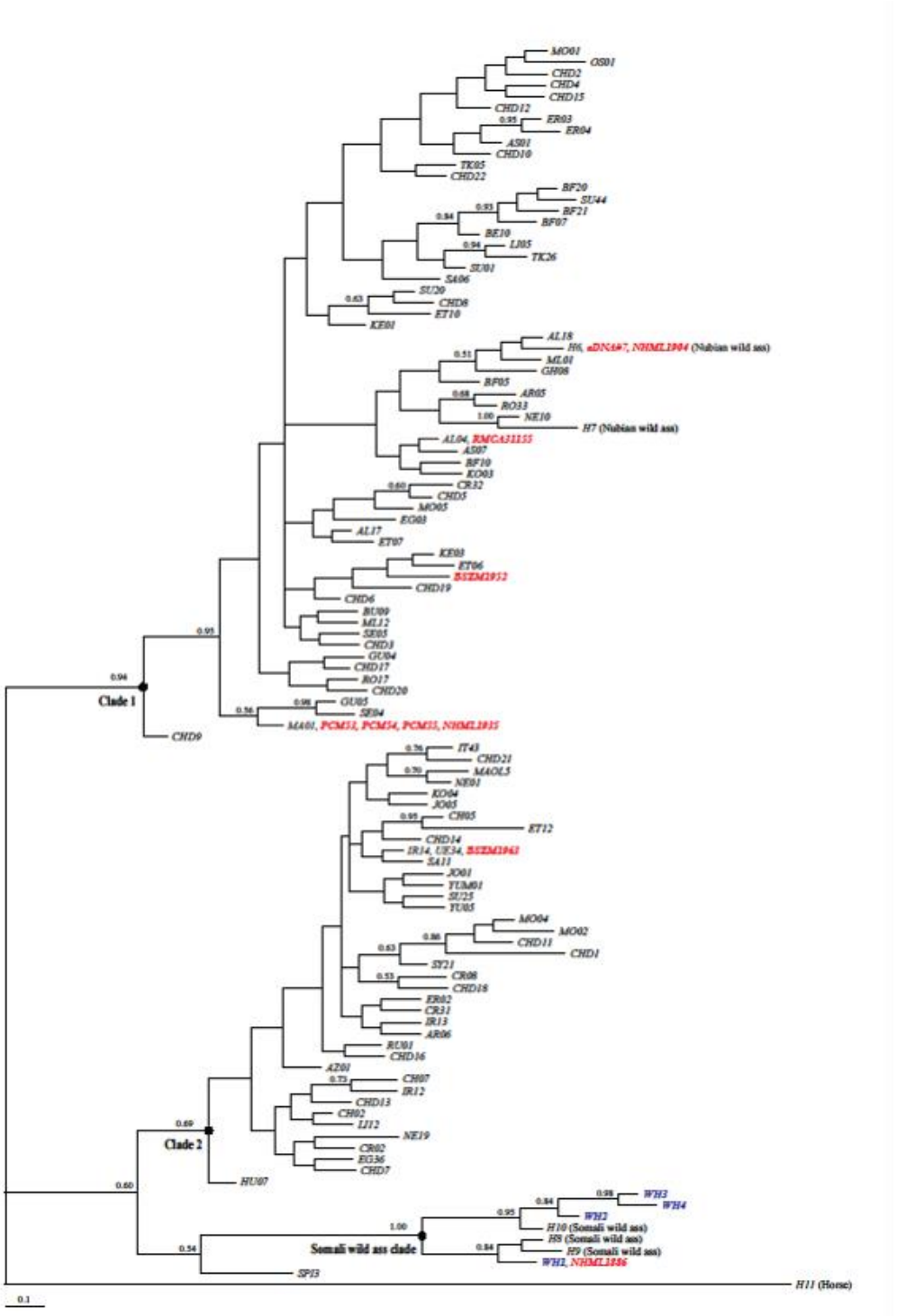


Figure S1 - Comparative morphometric analysis of the ancient Uan Muhaggiag magnum, Os carpale III with modern donkeys, Somali and Nubian wild ass and ancient donkeys from Abydos. GB = greatest breadth.



Figures S2 - Bayesian tree of 108 mtDNA haplotype sequences from domestic and wild asses, with horse sequence (H11, AY569548) as an outgroup. Numbers above nodes represent posterior probabilities (only values ≥ 0.50 are shown). Archaeological and historic sequences are marked in red color and fecal sequences are marked in blue.

Tables

Table S1 - Assayed ancient, historic, and modern samples

Ancient archaeological samples						
Sample ID	Origin	Age [a]	Collection [b]	Clade	Haplotype [c]	Sequence length (bps) [d]
aDNA#2 2005.1.209 #4 9999:0:1093 #5 9999:0:4324	Mugharet el Aliya Morocco	Early and later Holocene stratigraphic contexts	YPM	NA [e]		
aDNA#3 # 9999:0:4305	Grotte des Idoles Morocco	117 BC-AD 23 2082-2110 cal BP	YPM	Horse		
aDNA#6 Tu C5	Ti-n-Torah, Libya	c.7000 BC, c.10116 - 10248 cal BP	UR	NA [e]		
aDNA#7 [f] Verona 3870	Uan Muhaggiag Libya (trapezoid, Os carpale II)	1211-1026 cal BC 3160-2975 cal BP	CNHMV	1	H6	294 (3 amplicons)
aDNA#8 Verona 3988	Uan Muhaggiag Libya (mandible)	1211-1026 cal BC 3160-2975 cal BP	CNHMV	1	H6	201 (4 amplicons)
aDNA#14	Tarkan, Egypt	c.2000 BC (2)	NHML	NA [e]		
aDNA#15	Kerma, Sudan	c.1750 BC c.3800 cal BP	NHMG	NA [e]		
aDNA#1, NS1-F14	Narosura Kenya	c. 890 BC c.3800 cal BP	KNM	NA [e]		
aDNA#13,#16 BFIE 100	Tiha/-mah Yemen	c.4900 BC c.6800 cal BP	L. Khalidi NMS	NA [e]		
Historic museum samples						
PCM 53 Fetus *E.a.africanus*	Red Sea Hills Wadi Hafta 17°35' N37° 35'S	AD 1935	PCM	1	MA01	440
PCM 54 [g] Mother of PCM 55 *E.a.africanus*	Red Sea Hills Kilaiteit, Sudan 17°35'N 37° 35'S	AD 1935	PCM	1	MA01	440
PCM 55 Fetus *E.a.africanus*	Red Sea Hills Kilaiteit, Sudan	AD 1935	PCM	1	MA01	440
NHML1886 [g] Type Specimen *E. a. somaliensis*	Somalia near Berbera	AD 1886	NHML	Somali wild ass	WH1[h]	440
NHML1904 [g] *E .a .africanus*	Atbara, Nakheila N. Sudan	AD 1904	NHML	1	H6	440
NHML1935 [g] Mother of PCM 53 *E. a .africanus*	Red Sea Hills Wadi Hafta Sudan	AD1935	NHML	1	MA01	440
NHML1939 *E. a. africanus*	Atbara, Sudan	AD1939	NHML	1		205 (2 amplicons)
BSZM 1952 *E. a. africanus*	Eritrea	1930's	BSZM	1	BSZM1952 [h]	440
BSZM 1963 *E. a .africanus*	Eritrea	1930's	BSZM	2	UE34, IR14	440
RMCA 31155 *E. a. africanus*	Tibesti, Chad		RMCA	1	AL04	440
S001 *E. a. somaliensis*	Benadir, Somalia	AD 1930-1960	A. Bietti IPHR	NA [e]		
Modern samples						
WA1 (fecal)	Ethiopia 11º45.295'N 41º26.944'E	2002	F. Kebede	Somali wild ass	WH1[h]	440
WA2 (fecal)	Ethiopia 11º48.261'N 41º27.178'E	2002	F. Kebede	Somali wild ass	WH2[h]	440
WA4 (fecal)	Ethiopia 11º44.478'N 41º42.941'E	2005	F. Kebede	Somali wild ass	WH1[h]	440
WA5 (fecal)	Ethiopia 11º44.676'N 41º44.177'E	2005	F. Kebede	Somali wild ass	WH2[h]	440
WA6 (fecal)	Ethiopia 11º44.478'N 41º42.941'E	2005	F. Kebede	Somali wild ass	WH1[h]	440
WA7 (fecal)	Ethiopia 11º44.478'N 41º42.941'E	2005	F. Kebede	Somali wild ass	WH2[h]	440

WA11 (fecal)	Eritrea	2005	F. Kebede	Somali wild ass	WH2[h]	440
WA12 (fecal)	Eritrea	2005	F. Kebede	Somali wild ass	WH1[h]	440
WA13 (fecal)	Eritrea	2005	F. Kebede	Somali wild ass	WH1[h]	440
WA14 (fecal)	Eritrea	2006	F. Kebede	Somali wild ass	WH2[h]	440
WA15 (fecal)	Eritrea	2006	F. Kebede	Somali wild ass	WH2[h]	440
WA16 (fecal)	Eritrea 15°00.144N 40°02.267E	2007	F. Kebede	Somali wild ass	WH3[h]	440
WA17 (fecal)	Eritrea 15°00.308'N 40°02.218'E	2007	F. Kebede	Somali wild ass	WH2[h]	440
WA18 (fecal)	Eritrea	2007	F. Kebede	Somali wild ass	WH1[h]	440
WA19 (fecal)	Eritrea	2007	F. Kebede	Somali wild ass	WH2[h]	440
WA20 (fecal)	Eritrea	2007	F. Kebede	Somali wild ass	WH3[h]	440
WA21 (fecal)	Eritrea 15°01.464'N 40°01.107'E	2007	F. Kebede	Somali wild ass	WH4[h]	440
WA23 (fecal)	Eritrea 15°00.865'N 40°01.727'E	2007	F. Kebede	Somali wild ass	WH3[h]	440
WD1 (fecal)	Eritrea	2005	F. Kebede	Somali wild ass	WH2[h]	440
WD2 (fecal)	Eritrea	2005	F. Kebede	Somali wild ass	WH2[h]	440
WD3 (fecal)	Eritrea	2005	F. Kebede	Somali wild ass	WH2[h]	440
WD4 (fecal)	Eritrea	2005	F. Kebede	Somali wild ass	WH1[h]	440
WD6 (fecal)	Eritrea	2007	F. Kebede	Somali wild ass	WH3[h]	440
WD8 (fecal)	Eritrea	2007	F. Kebede	Somali wild ass	WH3[h]	440
WD9 (fecal)	Eritrea	2007	F. Kebede	Somali wild ass	WH1[h]	440
WD11 (fecal)	Eritrea	2007	F. Kebede	Somali wild ass	WH2[h]	440
WD12 (fecal)	Eritrea	2007	F. Kebede	Somali wild ass	WH4[h]	440
WD13 (fecal)	Eritrea	2007	F. Kebede	Somali wild ass	WH2[h]	440
WD14 (fecal)	Eritrea	2007	F. Kebede	Somali wild ass	WH3[h]	440
ErA2 (skin)	Asaila, Eritrea	2002	P. Moehlman	Somali wild ass	WH1[h]	440
ErA3 (skin)	Messir, Eritrea 14°59.963'N 40°00.622E	2006	P. Moehlman	Somali wild ass	WH3[h]	440
ErA4 (skin)	Asaila, Eritrea	2006	P. Moehlman	Somali wild ass	WH2[h]	440
ErA6 (skin)	Messir, Eritrea	2007	P. Moehlman	Somali wild ass	WH2[h]	440

[a] Direct dates of an unciform (Os carpale IV) associated with successfully amplified aDNA samples #7 and #8 are presented in Table S3. Direct dates for aDNA #14/Tarkan are reported by Burleigh (1). All other dates are relative stratigraphic estimates or based on associated radiocarbon dated material.

[b] Abbreviations used for institutions: BSZM = Bavarian State Collection for Zoology, Munich; CNHMV= Civic Natural History Museum, Verona; HE = Hoffman Ex.; IHPR = Institute of Human Palaeontology, Rome; KNM = Kenya National Museums, Nairobi; NHMG = Natural History Museum, Geneva; NHML = Natural History Museum, London; NMS = National Museum Sana'a; PCM= Powell-Cotton Museum, Kent; RMCA = Royal Museum of Central Africa, Tervuren; UR= University of Rome; YPM = Peabody Museum, Yale

University, New Haven. We are indebted to these institutions and individuals for research permission and assistance in collection of samples including, M. Harman, P. Jenkins, P. Kiura, R. Kraft and A. Vaccari.

[c] Haplotype designations are from Beja-Pereira *et al.* (2004).

[d] Total analyzed sequence length is shown for the samples that were successfully amplified and sequenced. Nucleotide positions of amplified fragments are as follows: aDNA sample #7 amplified bps 15408-15580, 15689-15774 and 15789-15823 (total of 294 bases) aDNA sample #8 amplified bps 15482-15580, 15689-15721, 15741-15773 and 15789-15823 (total of 201 bases); NHML 1939 amplified bps 15407-15526 and 15689-15773 (total of 205 bases); all other samples amplified bps 15407-15846 (440 bases).

[e] No amplification = NA

[f] Ancient sample verified at Harvard University.

[g] Historic samples verified at the CIBIO-Universidade do Porto

[h] New haplotypes from this study

Table S2 - Primer information

Primer	Sequence (3' to 5')	Annealing temperature (°C)	Product size (bp*)
d-loop			
L15386** with H15544	CCCAAGGACTATCAAGGAAG CTGATATGGTGTTGTGCG	55	158
L15462 with H15533	TGATTTCCTCCCCTAAACGA TTGTGCGGGGGTGTATTTTA	57	72
L15492 with H15600	CCCTCATGTGCYATGTCAGT GCRAACAATTTAATGCACGA	52	109
L15666 with H15865**	TACATACCCCATCCAAGTC GGAATGGCCCTGAAGAAAG	53	199
L15669 with H15724 or	ATACCCCATCCAAGTCAAAT GAATATGGGTGGTGATATGC	54	56
H15740 or	GGTGATTAAGCTCGTGGAA	54	72
H15791	GAGCGAGGATTGGGACAC	54	122
L15722 with H15791	TTCCACGAGCTTAATCACC GAGCGAGGATTGGGACAC	54	70
L15769 with H15844	ATTACGTGTCCCAATCCTC CCAGATGCCAGGTATAGTTTC	54	76
L15386 with H15694	CCCAAGGACTATCAAGGAAG GAAATGATTTGACTTGGATG	53	308
L15527 with H15744	CGCACAACACCATATCAG CTTGGTGATTAAGCTCGTGG	53	217
L15386** with H15865**	CCCAAGGACTATCAAGGAAG GGAATGGCCCTGAAGAAAG	55	479
M13			
M13longF M13longR	CACAGGAAACAGCTATGACCA TGTAAAACGACGGCCAGTG	57	Varied

* bp: base pairs including primers, ** Primers from Beja-Pereira *et al.* (2004)

Table S3 - Uan Muhaggiag and comparative magnum size data used in Figure 2

Identification	Specimen	Magnum GB	Magnum Max L	Age
unknown	Uan Muhuggiag	32	27	unknown
wild ass modern	NHMBA 10858, F	38	35	3-5 yrs
wild ass modern	FMNH 1427, M	39	34	3-5 yrs
wild ass modern	ZMB 70038, F	37	34	adult
wild ass modern	ZMB 30253, F	41	35	adult
wild ass modern	FMNH 18851, M	40	35	adult
wild ass modern	ZMB 46075, M	34	32	adult
wild ass modern	BSAPM 1, M	40	35	adult
wild ass modern	BSAPM 2, M	37	32	adult
wild ass modern	BSAPM 3, M	38	33	adult
wild ass modern	BSZM 1961.167, F	35	33	adult
wild ass modern	BSZM 1964.23, F	38	34	adult
wild ass modern	BSZM 1952.9, F	32	30	adult
wild ass modern	BSZM 1963.133, M	37	39	adult
wild ass modern	BSZM 1963.134, M	37	34	adult
wild ass modern	AMNH 135017, M	36		adult
donkey c. 3000 BC	Abydos 1, M	39	36	adult
donkey c. 3000 BC	Abydos IV, M	38	33	adult
donkey c. 3000 BC	Abydos V, M	36	34	adult
donkey c. 3000 BC	Abydos VI, M	37	33	adult
donkey c. 3000 BC	Abydos VIII	39	36	adult
donkey c. 3000 BC	Abydos IX	39	32	adult
donkey c. 3000 BC	Abydos X	36	32	adult
donkey, modern	KNM, OM5012, F	36	30	adult
donkey, modern	KNM, OM7061, M	34	32	adult
donkey, modern	BSAPM 4, F	31		adult
donkey, modern	BSAPM 5, F	34	31	adult
donkey, modern	BSZM 1968.696, F	33	29	adult
donkey, modern	NHMGE 825.1-2104.6, F	33	29	adult
donkey, modern	AMNH 204107, F	30		adult
donkey, modern	AMNH 15675, F	35		adult
donkey, modern	AMNH 100280, F	34		adult
wild ass modern	BSZM 1956.789, M	34		1.5 months

GB=greatest breadth following von den Driesch, Max L=maximum length,

Uan Muhaggiag mandible, dental age estimate 2.5-3 yrs, carpals may be derived from the same animal. Abbreviations used for institutions, additional to Table S1. AMNH=American Museum of Natural History, New York; FMNH=Field Museum of Natural History, Chicago; KNM=Kenya National Museums, Nairobi; NHMBA= Natural History Museum, Basel.

Table S4 - Mitochondrial DNA sequences of samples from current study

Position	15000+	411	484	489	490	503	536	539	541	569	579	580	598	599	621	636	637	644	652	Clade
	Sample ID																			
Reference	NC_001788	T	G	T	C	T	C	T	T	A	C	A	C	A	A	T	C	G	C	2
Nubian wild ass (historic)	BSZM1963	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	2
	NHML1935, PCM 53, 54, 55	-	A	-	T	C	-	-	-	G	-	G	T	G	G	-	-	A	T	1
	NHML1904	-	A	-	T	C	-	-	-	G	-	G	T	G	G	-	-	A	T	1
	BSZM1952	-	A	-	T	C	-	C	-	G	-	G	T	G	G	-	-	A	T	1
	RMCA31155	-	A	-	T	C	-	-	-	G	-	G	T	G	G	-	-	A	T	1
	NHML1939	-	A	-	T	C														1
Uan Muhaggiag (ancient)		-	A	-	T	C	-	-	-	G	-	G								1
Somali wild ass (modern and historic)	WA 2, 5, 7, 11, 14, 15, 17, 19, WD 1, 2, 3, 11, 13, ErA 4, 6	C	-	C	-	-	T	-	C	-	T	-	T	-	-	-	T	-	T	Somali wild ass
	WA 6, 20, 23, WD 6, 8, 14, ErA3	C	-	C	-	-	T	-	C	-	T	-	T	-	-	-	T	-	T	Somali wild ass
	WA 21, WD 12	C	-	C	-	-	T	-	C	-	T	-	T	-	-	-	C	-	T	Somali wild ass
	NHML1886, WA 1, 4, 6 12, 13, 18, WD 4, 9, ErA2	C	-	C	-	-	T	-	C	-	T	-	T	-	-	C	T	-	T	Somali wild ass
	15000+	662	667	698	704	707	713	714	718	770	771	801	802	803	806	820	821	822		
Reference	NC_001788	A	A	C	C	A	C	C	C	T	T	C	T	A	C	C	G	G		2
Nubian wild ass (historic)	BSZM1963	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		2
	NHML1935, PCM 53, 54, 55	G	-	T	-	-	-	-	-	C	-	T	-	-	T	T	A	-		1
	NHML1904	G	-	T	-	-	-	-	-	-	-	T	-	G	T	T	A	A		1
	BSZM1952	G	-	T	-	-	-	-	-	C	-	T	C	-	T	T	A	A		1
	RMCA31155	G	-	T	-	-	-	-	-	C	-	T	-	-	T	T	A	A		1
	NHML1939			T	-	-	-	-	-	-	-									1
Uan Muhaggiag (ancient)				T	-	-	-	-	-	-	-	T	-	G	T	T	A	A		1
Somali wild ass (modern and historic)	WH2	G	G	-	T	G	T	A	T	-	C	T	C	-	-	-	A	-		Somali wild ass
	WH3	G	G	T	T	G	T	A	-	-	C	T	C	-	-	-	A	-		Somali wild ass
	WH4	G	G	T	T	G	T	A	-	-	C	T	C	-	-	-	A	-		Somali wild ass
	WH1, NHML1886	-	G	-	T	-	T	A	T	-	-	-	-	-	-	-	A	-		Somali wild ass

Position numbers correspond to the reference sequence, i.e. NC_001788 (Xu *et al.* 1996).

Table S5 - Uan Muhaggiag radiocarbon dates

Laboratory Number	Material and Context	Radiocarbon Age	Calibrated Date Range (95% confidence interval)
Ox-17909 and Ox-17960 (*combined*)	*E. asinus* unciform (Os carpale IV) Pasa Trench1A, Level 2	2929 ± 20 bp	3160-2975 calBP 1211-1026 calBC
OxA-17909	*E. asinus* unciform:3870	2925 ± 27 bp	
OxA-17960	*E. asinus* unciform:3870	2915 ± 28 bp	

References

Baker S. (1867) *The Nile Tributaries of Abyssinia*. Macmillan, London.

Divé J. & Eisenmann V. (1991) Identification and discrimination of first phalanges from Pleistocene and modern Equus, wild and domestic. In: *Equids in the Ancient World* (ed. by Meadow RU, H. P.), pp. 278-333. Dr. Ludwig Reichert Wiesbaden.

Dollman G. (1935) A new race of wild ass from the Sudan, Asinus asinus dianae subsp. Nov. *Proceedings of the Linnean Society of London* **147**, 132-4.

Drummond A. & Rambaut A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**, 1-8.

Drummond A.J., Ho S.Y.W., Phillips M.J. & Rambaut A. (2006) Relaxed Phylogenetics and Dating with Confidence. *PLoS Biol* **4**, e88.

Groves C.P. (1986) The taxonomy, distribution, and adaptations of recent equids. In: *Equids in the ancient world* (eds. by Meadow RH & Uerpmann H-P), pp. 11-65, Wiesbaden.

Groves C.P. & Smeenk C. (2007) The nomeclature of the African wild ass. *Zoologische Mededelingen*, 121-35.

Kemp B.M., Malhi R.S., McDonough J., Bolnick D.A., Eshleman J.A., Rickards O., Martinez-Labarga C., Johnson J.R., Lorenz J.G., Dixon E.J., Fifield T.E., Heaton T.H., Worl R. & Smith D.G. (2007) Genetic analysis of early holocene skeletal remains from Alaska and its implications for the settlement of the Americas. *Am J Phys Anthropol* **132**, 605-21.

Lyddeker R. (1904) Notes on the specimens of wild asses in English collections *Novitates Zoologicae*, 583-96.

Lyddeker R. (1916) *Catalogue of the ungulate mammals in the British Museum (Natural History). Vo. V. Perissodactyla (horse, tapirs, rhinoceroses), Hyracoidea (hyraxes), Proboscidea (elephants)*. The Trustees of the British Museum, London.

Marshall F. & Weissbrod L. (2009) The consequences of women's use of donkeys for pastoral flexibility: Maasai ethnoarchaeology. In: *Tracking Down the Past. Ethnohistory Meets Archaeozoology* (eds. by Grupe G, McGlynn G & J. P), pp. 81-102. Verlag Marie Leidorf GmbH, Rhaden/Westf.

Murray G.W. (1935) *Sons of Ishmael*. Routledge, London, UK.

Nicolaisen J. (1963) *Ecology and culture of the Pastoral Tuareg*. The National Musum of Copenhagen, Copenhagen, Denmark.

Nicolaisen J. & Nicolaisen I. (1997) *The Pastoral Tuareg*. Thames and Hudson London.

Posada D. & Crandall K.A. (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817-8.

Rambaut A., Drummond A.J. (2007) Tracer version 1.4. URL http://beast.bio.ed.ac.uk/Tracer.

Richter W. (1982) *Lucius Iunius Moderatus Columella* Artemis Verlag, München, Zürich.

Rohland N. & Hofreiter M. (2007) Comparison and optimization of ancient DNA extraction. *Biotechniques* **42**, 343-52.

Ronquist F. & Huelsenbeck J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-4.

Rossel S., Marshall F., Peters J., Pilgram T., Adams M.D. & O'Connor D. (2008) Domestication of the donkey: timing, processes, and indicators. *Proc Natl Acad Sci U S A* **105**, 3715-20.

Rozen S. & Skaletsky H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**, 365-86.

Xu X., Gullberg A. & Arnason U. (1996) The complete mitochondrial DNA (mtDNA) of the donkey and mtDNA comparisons among four closely related mammalian species-pairs. *Journal of Molecular Evolution* **43**, 438-46.

Yalden D.W., Largen M.J. & Kock D. (1986) Catatalogue of the Mammals of Ethiopia 6. *Monitore Zoologico Italiano. Supplemento* **21**, 31-103.

Article 5

Animal Genetics, 46(1), 30-36

Genetic diversity in donkey populations from the putative centers of domestication

Sónia Rosenbom, Vânia Costa, Effa Kefena, Nasser Al-Araimi, Ahmed Abdel-Moneim, Mohammed A. Abdalla, Amel Bakhiet, Albano Beja-Pereira

Abstract

Donkey domestication has drastically changed ancient transport systems in Africa and Asia, enabling overland circulation of people and goods and influencing the organization of early cities and pastoral societies. Genetic studies based on mtDNA have pointed the African wild ass as the most probable ancestor of the domestic donkey, but questions regarding its center of origin remain unanswered. Endeavoring to pinpoint the geographical origin of domestic donkey, we have assessed levels and patterns of genetic diversity at 15 microsatellite loci from eight populations, representing its three hypothesized centers of origin: Northeast Africa, Near East and the Arabian Peninsula. Additionally, we have compared the donkey genotypes with those from their wild relative, the African wild ass (*Equus africanus africanus*), in order to visualize patterns of differentiation among wild and domestic individuals. Obtained results revealed limited variation in levels of unbiased expected heterozigozity (H_E) across populations in studied geographic regions (ranging from 0.637 in Northeast Africa to 0.679 in the Near East). Both allelic richness and private allelic richness presented considerable higher values in Northeast Africa and in the Arabian Peninsula. By looking at variation at country level, for each region, we were able to identify Sudan and Yemen as the countries possessing higher allelic richness and cumulatively Yemen presented also higher values for private allelic richness. Our results support previously proposed Northeast Africa as a putative center of origin, but the high levels of unique diversity in Yemen opens the possibility of considering this region as yet another center of origin for this species.

Introduction

Domestication of plants and animals has drastically transformed the course of human history. It provided early farmers with predictable food sources and influenced the density and mobility of human settlements, enabling the rise of civilization. Among earliest centers of domestication, the Fertile Crescent stands out as the most important source of livestock species, with goat, sheep, cattle and pig undergoing domestication events in a short period of time, between 11000 and 10000 years ago (Zeder 2008). This first wave of domestication provided the food resources for a rapid demographic expansion of human populations and triggered overland transport of people and goods. Increasing needs for efficient means of transportation lead to a second wave of domestication which is believed to have started in the Eurasian grasslands, about 6000 years ago, and from which have resulted the domestic horse, donkey and camel. The raise of the first organized conglomerates of people across Eurasia and Northern Africa and consequently the need of large amounts of supplies might have fomented the use of these species for transportation. Another reason is related with the climate changes suffered after the last maximum glaciation (around 18,000 years), from which resulted an increase of aridity in the Northern Africa and Southwest Asia. This fact might have propelled some herder communities to domesticate animals to help moving around the water sources and green pastures that have become more patchy and distant (Beja-Pereira *et al.* 2004; Rossel *et al.* 2008).

Donkey domestication has been a controversial theme, with zooarcheological, ethnographic and, more recently, genetic data, providing new insights into a complex scenario. Two alternative hypotheses arose for an African domestication of the donkey. The “Egyptian hypotheses” states that due to the presence of donkey bones in Predynastic Egyptian sites (6000-5000 BP), donkeys were most likely domesticated from resident Nubian wild ass (*E. africanus africanus*) by Egyptian villagers in the Nile Valley (Epstein 1971; Clutton-Brock 1992). A more recent hypothesis argues that donkey domestication occurred as a response of early pastoralists in Northeastern Africa to the increasing aridity in the Sahara (7000-6500 BP). The “pastoralist hypotheses” is well supported by ethnographic, climatic and linguistic data, and has become increasingly accepted (Marshall 2007). Besides this two African hypotheses for the domestication of the donkey, the identification of putative *Equus africanus* remains at sites in the Levant and the Arabian Peninsula, raised the possibility of a west Asian domestication of the donkey (Zeder 1986; Uerpmann 1987; Meadow R.H. 1991; Clutton-Brock 1992).

Genetic studies on donkey domestication have up to now relied on the analyses of mitochondrial DNA (mtDNA) variation from putative ancestors and contemporary domestic donkeys. The first

molecular study, by Beja-Pereira *et al.* (2004) identified the African wild ass as the probable ancestor of the domestic donkey, clearly ruling out the Asiatic wild ass as a putative progenitor. The existence of two clearly defined mtDNA clades among domestic donkeys suggested the occurrence of two independent domestication events, involving two distinct wild populations. Additionally, it was possible to identify one of the African wild ass subspecies - the Nubian wild ass - as the putative ancestor of Clade I donkeys. Ancestry of Clade II donkeys remains unknown, however it has been proposed that a relative of the Somali wild ass, probably already extinct, would be the most probable candidate (Beja-Pereira *et al.* 2004; Kimura *et al.* 2011). Besides northeastern Africa, both the ancient range of the Atlas wild ass in the Maghreb and the coast of Yemen remain potential geographical areas for the origin of the wild ancestor of Clade II donkeys (Kimura *et al.* 2011).

The African wild ass is critically endangered and facing a high risk of extinction (Moehlman 2008), however populations of the Somali wild ass (*Equus africanus somaliensis*) still subsist in Ethiopia and Eritrea. The Nubian wild ass (*Equus africanus africanus*), formerly distributed in Sudan and northern Eritrea, is currently very rare or even extinct.

Archeological data concerning the presence of donkeys in ancient societies is difficult to obtain because unlike cattle, donkeys were neither ceremonially buried nor a common subject in important art work (Marshall 2007). Nonetheless there are numerous sites containing donkey remains both in Africa (from the North African coast to the Horn) and in Asia (Arabian Peninsula and Near East). In fact, the largest known sample of *Equus africanus* or early *Equus asinus* from an archeological site is located in the Ash Shumah site in Yemen (Cattani & Bokonyi 2002), confirming the presence of African wild ass or an early domesticated form by 7770 ± 95 BP in the Arabian Peninsula.

Centers of origin are expected to retain more ancestral variation (Troy *et al.* 2001). As populations expand from centers of origin, genetic diversity is lost as a consequence of the limited numbers of individuals involved in this expansionist movements (“founder effect”). This pattern of decline in genetic diversity with increasing distance from proposed centers of origin has been found in cattle studies using autosomal markers (Loftus *et al.* 1999; Cymbron *et al.* 2005).

Hotspots of diversity, such as centers of origin, can often present a similar signature in terms of diversity as areas considered as melting pots (ancient trading areas or routes).

Distinguishing hotspots from melting pots can be done by analyzing allelic patterns and frequencies in order to discriminate shared diversity and unique diversity, typical in centers of

origin. This study stands as the first to assess levels of genetic variation among domestic donkeys from putative centers of origin and their wild counterpart (African wild ass), using nuclear markers.

Material and Methods

In the current study we have assessed levels of genetic diversity at 15 autosomal microsatellite loci, in donkey populations from 8 countries within the three hypothesized centers of origin of the domestic donkey: Northeastern Africa (Ethiopia, Sudan, Egypt), Arabian Peninsula (Oman, Yemen) and Near East (Syria, Turkey, Jordan) were sampled in order to assess levels and patterns of genetic diversity. Additionally, we have compared obtained domestic donkey genotypes with those obtained from captive Somali wild ass (N=20) (Rosenbom *et al.* 2012) at a subset of 10 microsatellites loci, in order to visualize patterns of differentiation among wild and domestic individuals.

DNA extraction and microsatellite genotyping

Blood and tissue samples belonging to donkeys from putative centers of origin were collected in a total of 129 (Table 1). DNA was extracted with the DNeasy Blood & Tissue kit (QIAGEN GmbH, Hilden), according to standard protocols. Samples were then diluted in elution buffer, according to the amount of DNA visible on the gel, and stored at -20ºC.

Genomic DNA was amplified by polymerase chain reaction (PCR), for 15 autosomal microsatellite loci (supplementary material - Table S1). Forward primers were modified by end labelling with fluorescent dyes (6-FAM™, VIC®, NED™, PET®) at the 5′ end. Each 15-µl reaction consisted of water, DNA, primers and fluorescent labels (0.06 µM primer forward, 0.6 µM primer reverse and dye), dNTPs(30 mM each), 10× buffer [200 mM Tris–HCl (pH 8.4), 500 mM KCl], BSA (0.4 µg/µl), MgCl2 (variable between 1.5 and 3 mM, according to the loci) and Platinum® Taq DNA Polymerase ((0.3 U); Invitrogen™). Samples were amplified in a Dual 96-Well GeneAmp® PCR System 9700 thermocycler (Applied Biosystems™) in the following conditions: initial denaturation at 95°C for 10 min, followed by 35 cycles of 30 s at 95°C, 30 s at temperatures between 52°C and 60°C (variable with the loci) and 30 s at 72°C; a final elongation step was held for 10 min at 72°C. PCR products were checked in 2% agarose gel stained with GelRed™ and, according to the quality of the amplification, diluted in water, mixed with formamide and LIZ® 500-bp internal size standard (Applied Biosystems™) and detected by capillary electrophoresis using a 3100

Genetic Analyzer® (Applied Biosystems™) sequencer. Software GeneMapper® v4.0 (Applied Biosystems™) was used to score individual genotypes.

Statistical analyses

Genetic diversity

Deviations from Hardy–Weinberg equilibrium proportions were tested, using program GENEPOP v4.2. (Raymond & Rousset 1995), for each population–locus combination. Departure from Hardy–Weinberg expectations was assessed by exact tests with unbiased P values estimated using a Markov chain method (set to 1000 batches of 10 000 iterations each and with 10 000 steps of dememorization). A global test across loci and populations was performed using the Fisher's method. The null hypothesis of no genotypic linkage disequilibrium was tested between all pairs of loci in each population and additionally a global test (Fisher's method) for each pair of loci was performed across samples. Statistical significance was adjusted for multiple comparisons by using sequential Bonferroni correction (Rice 1989).

For geographic comparisons donkey samples were grouped in three regions corresponding to three putative centers of origin - Near East, Arabian Peninsula and Northeast Africa. Levels of genetic diversity were assessed by calculating unbiased expected (uH_E), expected (H_E) and observed heterozygosities (H_O) per each proposed geographical region using GenAlEx 6.5 software (Peakall & Smouse 2006; Peakall & Smouse 2012).

Allelic richness (Ar) as well as private allelic richness (PAr) were estimated, per each geographic region, by using the rarefaction algorithm implemented in HP-RARE 1.0 (Kalinowski 2005), in order to account for potential biases arising from unequal sample sizes. Estimates of these two measures were standardized to the smallest sample size (Near East; N=20).

Statistical significance of obtained differences for calculated diversity measures (uH_E, Ar and PAr) was accessed by conducting Mann-Whitney tests as implemented in MINITAB® Statistical Software.

Levels of genetic diversity were further investigated, by zooming into individual countries, in regions corresponding to putative centers of origin (Near East, Arabian Peninsula and Northeast Africa), by calculating above mentioned diversity parameters, using GenAlEx 6.5 software (Peakall & Smouse 2006; Peakall & Smouse 2012). We have also used HP-RARE 1.0 (Kalinowski

2005) in order to estimate allelic richness (Ar) as well as private allelic richness (PAr), per country, correcting these values for the smallest sample size (Egypt; N=15).

Allelic patterns

Allelic frequencies and distributions, per locus, were analyzed across geographic regions using GenAlEx 6.5 software (Peakall & Smouse 2006; Peakall & Smouse 2012). A Mann-Whitney test was conducted to test if differences in allelic frequencies among regions were statistically significant, using MINITAB® Statistical Software. In order to distinguish unique from shared diversity we looked at putative private alleles per geographic region, with frequencies between 2 and 5%, in order to exclude rare alleles caused by random sampling. Additionally, we have looked for divergent alleles in the frequency distribution plot, per locus, in order to identify their geographical origin.

Population differentiation

Analysis of molecular variance (AMOVA) (Excoffier *et al.* 1992) was performed using GenAlEx 6.5 software (Peakall & Smouse 2006; Peakall & Smouse 2012). Weir & Cockerham (1984) hierarquical *F*-statistics were used to estimate the proportion of genetic variability found among domestic donkey populations (F_{ST}), among individuals belonging to the same population (F_{IS}) and within individuals (F_{IT}). To determine if observed level of differentiation was significantly greater than that expected by chance, we compared the obtained value, against the outcomes of 1000 permutations.

In order to visualize similarities and dissimilarities among sampled donkey populations belonging to countries of interest (Egypt, Ethiopia, Sudan, Yemen, Oman, Syria and Turkey) and between these and their wild counterpart (African wild ass), a factorial correspondence analyses (FCA) was performed using GENETIX v. 4.05 (Belkhir K. 1996-2004).

Results

Genetic diversity

Two locus-population combinations significantly ($P<0.001$) deviated from Hardy-Weinberg expectations (locus VHL20 in Sudan and locus CA425 in Yemen), however no locus significantly

deviated from HWE proportions across populations. Results showed no significant gametic (linkage) disequilibrium between all possible loci pairs, across populations, after Bonferroni corrections for multiple tests.

A total of 120 alleles were detected in the 15 surveyed loci, giving a mean number of 8 alleles per locus. Obtained heterozygosity values (both unbiased expected and observed) showed some variation across regions. Unbiased expected (uH_E) heterozygosity ranged from 0.637 ± 0.056 in Northeast Africa to 0.679 ± 0.049 in the Near East (Table 1), however values for observed heterozygisty (H_O) were higher in the Arabian Peninsula region (0.613 ± 0.048). Patterns of diversity given by allelic richness (Ar) and private allelic richness (PAr) were somewhat different. Values of both allelic richness and private allelic richness were higher in the Arabian Peninsula (Ar=5.93; Par=0.65) and Northeast Africa (Ar=5.81; Par=0.53) regions, what might be indicative of a source of unique, unshared, genetic diversity, consistent with what is expected in centers of origin. Results for Mann-Whitney tests showed no support for significant differences ($P<0.05$) in diversity parameters (uH_E, Ar and PAr), among geographic regions.

Table 1 - Summary statistics of sampled domestic donkeys. Represented parameters are mean values per geographic region, across loci. Number of samples is indicated (N) as well as observed heterozygosity (HO), expected heterozygosity (HE), unbiased expected heterozygosity (uHE), allelic richness (Ar) and private allelic richness (PAr). Allelic richness (Ar) and private allelic richness (PAr) were calculated using the rarefaction algorithm for the minimum sample size in the Near East region (N=20)

Region	N	H_O	H_E	uH_E	Ar (N=20)	PAr (N=20)
Northeast Africa	60	0.579 ± 0.055	0.631 ± 0.055	0.637 ± 0.056	5.81	0.53
Near East	20	0.564 ± 0.052	0.658 ± 0.047	0.679 ± 0.049	5.67	0.27
Arabian Peninsula	49	0.613 ± 0.048	0.656 ± 0.041	0.665 ± 0.041	5.93	0.65

Northeast Africa *- Egypt, Sudan, Ethiopia;* ***Near East*** *- Turkey, Syria, Jordan;* ***Arabian Peninsula*** *- Oman, Yemen*

Genetic diversity was further investigated by zooming into countries in regions of interest. Obtained results showed Sudan as possessing the highest levels for calculated genetic diversity parameters (H_O, H_E, uH_E and Ar) (Table 2). Yemen possessed the highest values for private allelic

richness (PAr = 0.55) and the second highest values for allelic richness (Ar = 5.78). Sudan and Yemen stand out, among countries in putative centers of origin, as those possessing the highest values for analyzed diversity parameters, indicating these countries in particular as potential sources of genetic diversity.

Table 2 - Summary statistics of domestic donkeys from countries in putative centres of origin. Represented parameters are mean values, across loci. Number of samples is indicated (N) as well as allelic richness (Ar), observed heterozygosity (H_O), expected heterozygosity (H_E) and unbiased expected heterozygosity (uH_E)

Country	N	H_O	H_E	UH_E	Ar(N=15)	Par(N=15)
Egypt	15	0.538 ± 0.06	0.537 ± 0.058	0.557 ± 0.059	4.78	0.12
Turkye+Syria	17	0.555 ± 0.06	0.58 ± 0.057	0.6 ± 0.058	4.64	0.12
Oman	20	0.590 ± 0.064	0.622 ± 0.048	0.648 ± 0.05	5.26	0.28
Sudan	20	0.647 ± 0.054	0.64 ± 0.051	0.667 ± 0.053	6.07	0.29
Yemen	29	0.565 ± 0.066	0.598 ± 0.062	0.61 ± 0.063	5.78	0.55
Ethiopia	25	0.464 ± 0.075	0.486 ± 0.077	0.496 ± 0.079	5.26	0.27

Allelic patterns

Analyzes of allelic patterns across geographic regions revealed the existence of 24 private alleles across regions, being that 96% of those alleles were found in the Northeast Africa and Arabian Peninsula regions. However, when looking at allelic frequencies, approximately 64% presented frequencies between 2% and 5% in the Arabian Peninsula, while only one private allele in the Northeast Africa region showed frequencies above 2%. Additionally, we have observed the relative position of these private alleles in the frequency plots, per locus. Obtained results showed that private alleles identified in the Arabian Peninsula at frequencies between 2 and 5% were also the most divergent, presenting a marginal distribution in the frequency plots (Fig. 1). However, when testing for significant differences between allele frequencies distributions across regions, using Mann-Whitney tests, no significant differences were found ($P<0.05$).

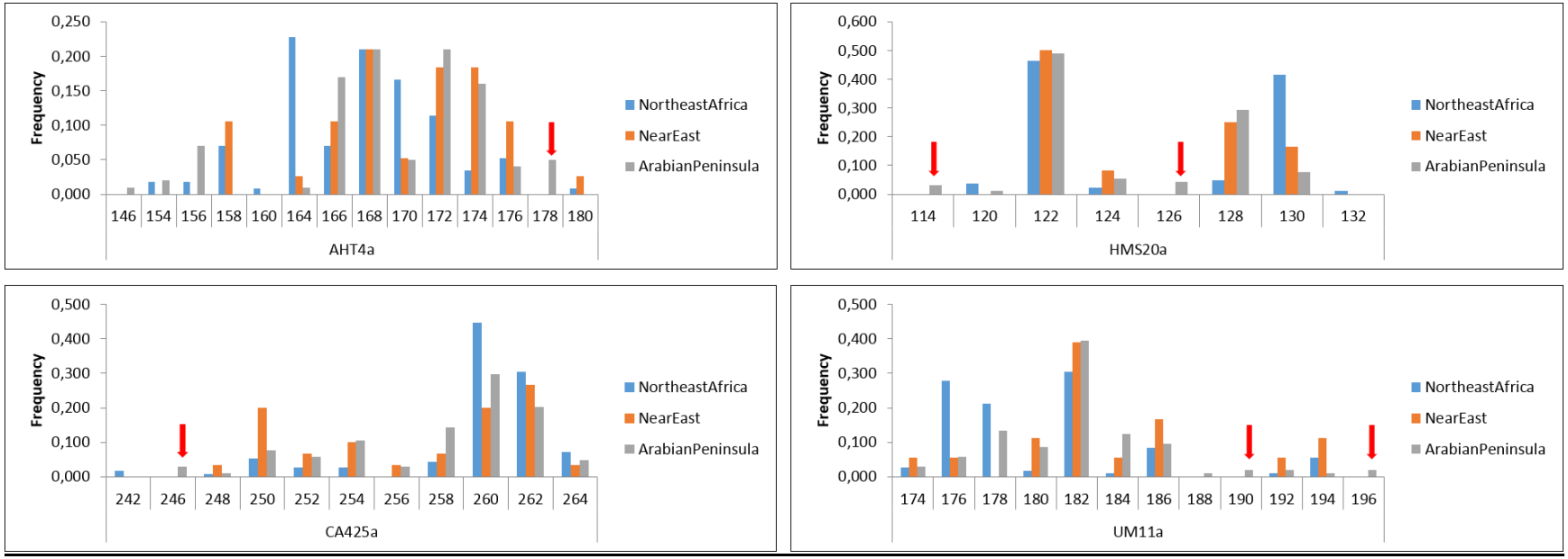


Figure 1 - Allelic frequencies distribution across regions for the four loci showing a large number of private alleles identified at frequencies between 2 and 5% (AHT4, HMS20, CA425 and UM11). Red arrows denote private alleles.

Population differentiation

Analyses of molecular variance (AMOVA) among domestic donkey populations revealed low differentiation, with only 10% of found variation justifying differences among populations (F_{ST}) and 22% of obtained variation being justified by differences among individuals belonging to the same population (F_{IS}). Variation within individuals (F_{IT}) was responsible for approximately 68% of obtained genetic diversity.

Pairwise F_{ST} values among domestic populations from countries in putative centers of origin and the African wild ass varied between 13% and 17% (for the Sudanese and Egyptian populations respectively). Pairwise F_{ST} between the Yemeni donkey population and the African wild ass population was the second lowest (13.6%), after Sudan (supplementary material - Table 2). Also the factorial components analyses (FCA) revealed clear differentiation among domestic donkeys and their wild counterpart (Fig. 2). Obtained pairwise F_{ST} values among Yemeni donkeys and the remaining domestic populations were also the highest among all possible domestic population pairs (supplementary material - Table 2).

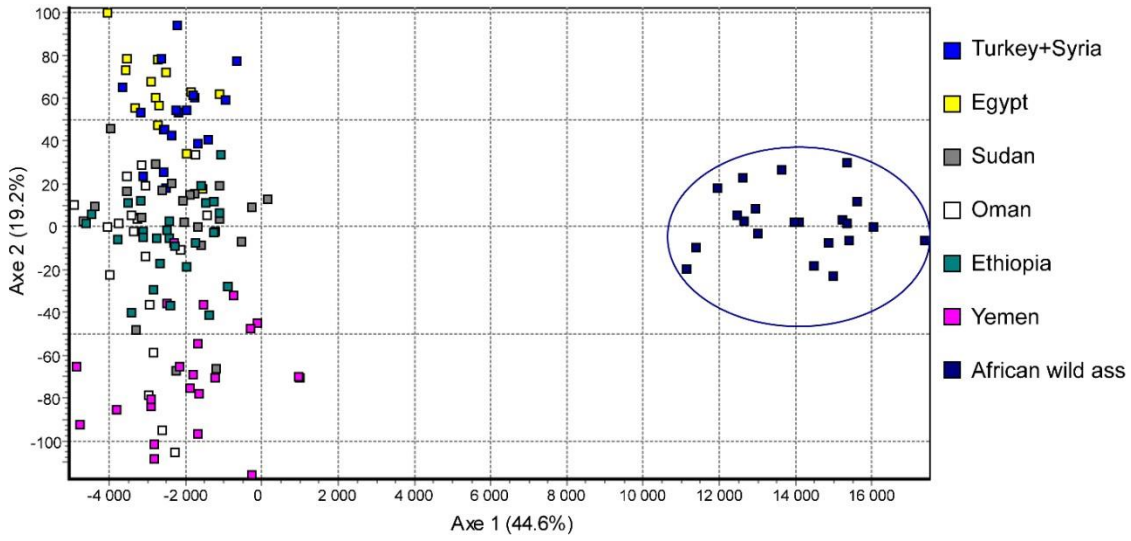


Figure 2 - Factorial correspondence analyses (FCA) of domestic donkey populations belonging to putative centers of origin and an African wild ass (*Equus africanus*) population. African wild ass individuals are identified in the blue circle.

Discussion

Donkey domestication has proved to be a complex process, with different putative ancestors being proposed along the course of history, as well as different geographic regions being

pinpointed as potential areas for donkey domestication, on the basis of historical and archeological data. Genetic studies clearly pointed the African wild ass as the ancestor of the domestic donkey, in particular the Nubian wild ass (*Equus africanus africanus*) was identified as the putative ancestor of Clade I donkeys (Beja-Pereira *et al.* 2004; Kimura *et al.* 2011).

Despite being currently critically endangered or even extinct in the wild, the Nubian Wild Ass was historically distributed in Sudan and northern Eritrea (Moehlman 2008). Based on the geographical distribution of this subspecies and its role in domestic donkey ancestry, our hypotheses was that domestic populations occurring in close proximity with geographic areas where the Nubian wild ass was distributed would present higher levels of genetic diversity. Obtained results clearly pinpointed Sudan as presenting the highest values of unbiased expected heterozigosity, as well as allelic richness, supporting our initial hypothesis. Levels of diversity in other analyzed countries in Northeast Africa, namely in Egypt and Ethiopia, were considerably lower for all diversity parameters (Table 2). On the light of newly obtained genetic diversity data, the alternative pastoralist hypothesis for donkey domestication in Northeast Africa is favored in detriment of an Egyptian domestication.

Zooming into individual countries, Yemen stands out as the population presenting higher values of private allelic richness, pinpointing this country as a potential source of unique, unshared diversity among studied populations. The presence of African wild ass or early domesticated donkeys in a site in Yemen is well-documented and considered reliable among archaeologists (Marshall 2007). Dating of excavated bones, at about 7770 ± 95 BP predates the proposed date for donkey domestication at approximately 5000 years ago (Beja-Pereira *et al.* 2004; Kimura *et al.* 2011), making plausible the existence of a wild population, in the Arabian Peninsula, during ancient times. This scenario, favors a possible domestication event in the Arabian Peninsula, from an already extinct wild ass population, as suggested by archaeological data (Cattani & Bokonyi 2002).

The absence of geographical structure among studied domestic donkey populations, using a set of nuclear markers, is consistent with previously reported absence of structure in mitochondrial DNA donkey haplotypes (Beja-Pereira *et al.* 2004; Pérez-Pardal *et al.* 2014). This pattern can be explained by some specificity of the donkey and its domestication process, namely the mobile character of the species and the absence of intensive management (artificial selection) in most regions of the word, unlike the closely related horse. Despite the short time span since domestication, domestic donkeys and the African wild ass can clearly be differentiated.

Conclusion

As the first study using nuclear markers and samples belonging to domestic donkey populations from putative centers of origin and their wild counterpart (African wild ass), the obtained results point to two geographic regions as possessing the highest genetic diversity values – Northeast Africa and the Arabian Peninsula. Also at a finer scale, Sudan and Yemen are the two countries where the highest values of diversity were displayed. Although these results confirm previously reported studies on mtDNA which pointed towards Northeast Africa as the potential center of origin of the donkey, the high diversity levels found in the Arabian Peninsula, namely in Yemen, are not negligible and suggest a possible involvement of this region in donkey domestication. Such similar high-diversity levels also point towards a much more complex domestication process where wild animals from several geographically distinct populations might have been recruited. This and the fact that the domestic donkey purpose was the transportation of goods and people, led to different gene pools getting swiftly mixed and distributed all over the ancient world.

Acknowledgments

This work was funded by FEDER funds though a project co-funded by the EU program COMPETE and the national funds from the Portuguese Foundation for Science and Technology (FCT) (PTDC/BIA-BIC/118107/2010 and FCOMP-01-0124-FEDER-019757) and from another FCT project (PTDC/BIA-BDE/64111/2006). AB-P is supported by FCT through an IF-FCT contract.

Conflict of interests:

The authors declare no conflict of interests to declare.

References

Beja-Pereira A., England P.R., Ferrand N., Jordan S., Bakhiet A.O., Abdalla M.A., Mashkour M., Jordana J., Taberlet P. & Luikart G. (2004) African origins of the domestic donkey. *Science* **304**, 1781.

Belkhir K. B.P., Chikhi L., Raufaste N. & Bonhomme F. (1996-2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II. Montpellier (France).

Cattani M. & Bokonyi S. (2002) Ash-Shumah: an early Holocene settlement of desert hunters and mangrove foragers in the Yemeni Tihamah In: *Essay on the late Prehistory of the Arabian*

Peninsula (eds. by Cleuziou S, Tosi M & Zarins J), pp. 31-53. Instituto Italiano per l'Afirca e l'Oriente, Rome.

Clutton-Brock J. (1992) *Horse power: a history of the horse and the donkey in human societies*. Harvard University Press.

Cymbron T., Freeman A.R., Isabel Malheiro M., Vigne J.-D. & Bradley D.G. (2005) Microsatellite diversity suggests different histories for Mediterranean and Northern European cattle populations. *Proceedings of the Royal Society B: Biological Sciences* **272**, 1837-43.

Epstein H. (1971) *The origin of the domestic animals of Africa*. African Publishing Corporation (APC), 101 Fifth Avenue, New York, N.Y. 100003, USA.

Excoffier L., Smouse P.E. & Quattro J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-91.

Kalinowski S.T. (2005) hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* **5**, 187-9.

Kimura B., Marshall F.B., Chen S., Rosenbom S., Moehlman P.D., Tuross N., Sabin R.C., Peters J., Barich B., Yohannes H., Kebede F., Teclai R., Beja-Pereira A. & Mulligan C.J. (2011) Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication. *Proceedings of the Royal Society B: Biological Sciences* **278**, 50-7.

Loftus R.T., Ertugrul O., Harba A.H., El-Barody M.A.A., MacHugh D.E., Park S.D.E. & Bradley D.G. (1999) A microsatellite survey of cattle from a centre of origin: the Near East. *Molecular Ecology* **8**, 2015-22.

Marshall F. (2007) *Rethinking Agriculture: Archeological and Ethnoarcheological perspectives*. Left Coast Press, Inc., Walnut Creek, California, USA.

Meadow R.H. U.H.-P. (1991) *Equids in the Ancient World*. Ludwig Reichert Verlag, Wiesbaden.

Moehlman P.D., Yohannes, H., Teclai, R. & Kebede, F. (2008) *Equus africanus*. URL http:\\www.iucnredlist.org

Peakall R. & Smouse P.E. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28**, 2537-9.

Peakall R.O.D. & Smouse P.E. (2006) genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-95.

Pérez-Pardal L., Grizelj J., Traoré A., Cubric-Curik V., Arsenos G., Dovenski T., Marković B., Fernández I., Cuervo M., Álvarez I., Beja-Pereira A., Curik I. & Goyache F. (2014) Lack of mitochondrial DNA structure in Balkan donkey is consistent with a quick spread of the species after domestication. *Animal Genetics* **45**, 144-7.

Raymond M. & Rousset F. (1995) GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. *Journal of Heredity* **86**, 248-9.

Rice W.R. (1989) Analyzing tables of statistical tests. *Evolution*, 223-5.

Rosenbom S., Costa V., Steck B., Moehlman P. & Beja-Pereira A. (2012) Cross-species genetic markers: a useful tool to study the world's most threatened wild equid—Equus africanus. *European Journal of Wildlife Research* **58**, 609-13.

Rossel S., Marshall F., Peters J., Pilgram T., Adams M.D. & O'Connor D. (2008) Domestication of the donkey: timing, processes, and indicators. *Proc Natl Acad Sci U S A* **105**, 3715-20.

Troy C.S., MacHugh D.E., Bailey J.F., Magee D.A., Loftus R.T., Cunningham P., Chamberlain A.T., Sykes B.C. & Bradley D.G. (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature* **410**, 1088-91.

Uerpmann H.-P. (1987) *The Ancient Distribution of Ungulates Mammals in the Middle East* Ludwig Reichert Verlag, Wiesbaden.

Zeder M.A. (1986) The equid remains from Tal-e-Malyan, southern Iran. In: *Equids in the Ancient World* (ed. by Meadow RH UH-P). Ludwig Reichert Verlag Wiesbaden.

Zeder M.A. (2008) Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proceedings of the National Academy of Sciences* **105**, 11597-604.

Supplementary material

Table S1 – Primer sequences, dye labels and bibliographic reference for the 15 amplified microsatellite loci

Locus	Primer seqences	Dye label	Reference
AHT4	AACCGCCTGAGCAAGGAAGT GCTCCCAGAGAGTTTACCCT	6FAM	Binns et al. 1995
COR20	TCTCTACCGCAAGTGAAACC CTGAATTGTAGGACATCCCG	NED	Hopman et al. 1999
COR58	GGGAAGGACGATGAGTGAC CACCAGGCTAAGTAGCCAAAG	6FAM	Ruth et al. 1999
COR90	GGTTTGTCTCTTTGAGGTGTG TGCTCATATCTTCACCCTGC	NED	Tallmadge et al. 1999
HTG6	TAATACGACTCACTATAGG GTTCACTGAATGTCAAATTCTGCT	VIC	Marklund et al. 1994
HMS20	TGGGAGAGGTACCTGAAATGTAC GTTGCTATAAAAAATTGTCTCCCTAC	6FAM	Guérin and Bertaud 1996
HMS7	CAGGAAACTCATGTTGATACCATC TGTTGTTGAAACATACCTTGACTGT	6FAM	Guérin et al. 1994
VHL20	CAAGTCCTCTTACTTGAAGACTAG AACTCAGGGAGAATCTTCCTGAG	6FAM	van Haeringen et al. 1994
HMS6	GAAGCTGCCAGTATTCAACCATTG CTCCATCTTGTGAAGTGTAACTCA	VIC	Guérin et al. 1994
CA425	AGCTGCCTCGTTAATTCA CTCATGTCCGCTTGTCTC	PET	Eggleston-Stott et al. 1999
UM11	TGAAAGTAGAAAGGGATGTGG TCTCAGAGCAGAAGTCCCTG	NED	Meyer et al. 1997
COR32	GCCCTCTTAGAGCATTTTCC CAGAGATGGCTGGAGTAAGG	6FAM	Murphie et al 1999
LEX74	AAGAGTGCTCCCGTGTG GACAATGCAGAACTGGGTAA	PET	Bailey et al. 2000
NVHEQ18	GGAGGAGACAGTGGCCCCAGTC GCTGAGCTCTCCCATCCCATCG	VIC	Røed et al. 1997
AHT5	ACGGACACATCCCTGCCTGC GCAGGCTAAGGGGGCTCAGC	VIC	Binns et al. 1995

Table S2 – Pairwise F_{ST} distances among donkey populations from countries of interest and the African wild ass (Somali wild ass)

Egypt	Turkey+Syria	Oman	Sudan	Yemen	Ethiopia	Somali wild ass	
0.000							**Egypt**
0.033	0.000						**Turkey+Syria**
0.066	0.064	0.000					**Oman**
0.044	0.047	0.024	0.000				**Sudan**
0.141	0.115	0.083	0.091	0.000			**Yemen**
0.063	0.067	0.064	0.045	0.088	0.000		**Ethiopia**
0.173	0.159	0.155	0.131	0.136	0.140	0.000	**Somali wild ass**

Chapter V

Ancillary work

Article 6

Mitochondrial DNA, 2014 (doi: 10.3109/19401736.2014.898276)

Genetic diversity of the Ethiopian Grevy's zebra populations that includes a unique population of the Alledeghi Plain

Fanuel Kebede, **Sonia Rosenbom**, Leili Khalatbari, Patricia D. Moehlman, Albano Beja-Pereira, and Afework Bekele

Abstract

The endangered Grevy's Zebra *(Equus grevyi*) is confined to the Horn of Africa, specifically Ethiopia and Kenya. It is threatened by habitat loss and fragmentation due to human encroachment of historic range. Knowledge of population genetics is essential for the development of appropriate conservation actions and management. The focus of this study was to assess the heterogeneity and genetic distinctiveness of the two Grevy's zebra populations in Ethiopia. Non-invasive fecal samples (N = 120) were collected during 2009–2010 from Grevy's zebra populations in the Alledeghi Wildlife Reserve and the Sarite area, Ethiopia.

Analyses of a 329 bp of the mtDNA control region of 47 sequences, revealed the existence of two unreported haplotypes in the northern population of Alledeghi, that were not shared with the southern population of Sarite. The Sarite population is contiguous with the Grevy's zebra population in Kenya. The nucleotide diversity levels found in both the populations are extremely low.

Introduction

Historically, the Grevy's zebra ranged from the Alledeghi Plain in northern Ethiopia, through the Awash River valley to the Ogaden to southwestern Somalia, to the northeast of Lake Turkana and south into northern Kenya (Bauer *et al.* 1994). Currently the Grevy's zebra has a discontinuous range (Fig. 1).

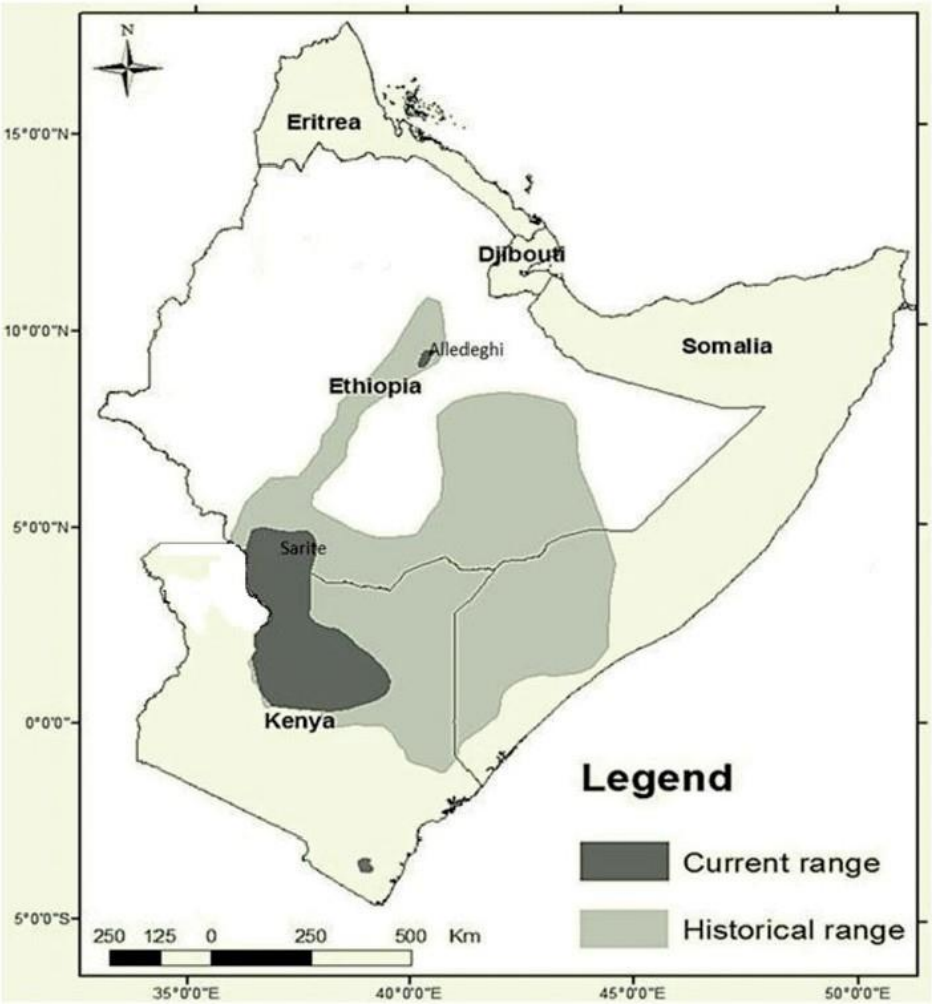


Figure 1 - Historic and current distribution of Grevy's zebra (based on Bauer et al., 1994).

There is a small isolated population in the Alledeghi Plain in northern Ethiopia. In southern Ethiopia (≈ 700 km from Alledeghi), there is a population that extends from Sarite/Chew Bahir into northern Kenya to Laikipia near Mt. Kenya (Fig. 1).

From 1980 to 2007, the global population of Grevy's Zebra declined by 68%. The current population is estimated to be approximately 2700 individuals (April 2012 Kenya Wildlife Service workshop), and this species is listed as Endangered by the IUCN Red List (Moehlman *et al.* 2008). In Ethiopia, the population decline was severe and went from an estimated 1900 in 1980 to approximately 128 in 2006 (Moehlman *et al.* 2008). The available suitable habitat has been dramatically reduced in Ethiopia due to increasing human populations, development of

rangelands for agriculture and increasing competition for resources, particularly water, with local pastoral people and their livestock (Kebede *et al.* 2012).

Grevy's zebra is listed as Endangered on the IUCN Red List and is CITES Appendix I. However, poaching for meat and medicine continues to be a threat Grevy's zebra survival (Williams 2002; Kebede *et al.* 2012). In recent years, the global population has spread farther south in Kenya to the Laikipia Plateau which was not part of the historic range. This range shift resulted in a larger overlap in populations of Grevy's and Plains zebra and adds one more source of concern, the increased potential for hybridization in the Kenyan portion of the Grevy's zebra population. Hybridization between these species could compromise their genetic integrity (Cordingley *et al.* 2009). The isolated population of Grevy's zebra in the Alledeghi Plain of Ethiopia is sheltered from this problem, but still faces challenges such as inbreeding and genetic drift.

Basic knowledge of levels of genetic diversity in endangered species and their disparate populations is an essential step for determining conservation actions that are sensitive to units of conservation and appropriate management plans for the entire range of Grevy's zebra (Moritz 1994; Frankham 2002).

Collection of samples from natural populations using a noninvasive method is a great advantage when studying elusive, rare and/or endangered species [e.g. Beja-Pereira *et al.* (2009)]. However, there can be problems with analyzing such samples, i.e. fecal, due to the poor quality/quantity of the obtained DNA (Taberlet *et al.* 1996; Broquet *et al.* 2007).

We have investigated levels of genetic diversity in two populations of Grevy's zebra in Ethiopia using sequence analyses of a mtDNA control region fragment. We discuss the results in light of previous studies reporting levels and patterns of nucleotide (π) and haplotype (h) diversity in populations of Plains zebra (*Equus quagga*) (Lorenzen *et al.* 2008) and Mountain zebra (*Equus zebra*) (Moodley & Harley 2005) and provide important first insights into genetic diversity of this endangered African equid.

Materials and Methods

A total of 120 noninvasive samples were collected from two sites in Ethiopia, i.e. Alledeghi (n = 86) and Sarite (n = 34) during the 2009 and 2010 wildlife surveys. Fecal samples were collected in the field and stored in individual brown paper bags with details on the collection site, time and date. The samples were dried naturally and stored at room temperature until further processing. A total of 82 individuals were selected for DNA extraction. In order to minimize potential

contamination issues inherent to non-invasive samples, DNA extraction was carried on a laminar flux chamber, physically separated from the PCR room. The samples were processed in batches with a maximum of 16 samples per set. All the materials used for the extraction process was disinfected between samples processing and between all groups of samples as well. In each batch of samples, a negative control containing all reagents but not the sample was included to detect contaminations. The fecal samples were processed for DNA extraction using JETquick Tissue DNA Spin Column (Genomed; Bad Oeynhausen, Germany). For each sample, the external part, containing the intestinal epithelial cells, was removed to a falcon tube. Standard lysis buffer and proteinase K were added and the samples were digested overnight at 56°C. After digestion, the samples were centrifuged and the supernatant was removed. All the solid components were discarded and the samples supernatant were processed as blood following the extraction kit manufacturer's instructions until the elution step. In this step, two elutions were carried out, each one for a different tube resulting in two replicates at the end. After the DNA extraction, both replicates of all samples were tested on 0.8% agarose gel and visualized in order to test the success of the extraction. According to the amount of DNA visible on the gel, the samples were diluted in buffer and stored at -20°C until the PCR. Approximately 2–5 ng of genomic DNA was used as template to amplify. A 350 bp fragment of the mitochondrial control region was amplified using the primers Eq-CR-1F (CCTCATGTACTATGTCAGTA) and Eq-Cr-534R (CCTGAAGAAAGAACCAGATGCC). Samples were amplified in a Dual 96-Well GeneAmp® PCR System 9700 thermocycler (Applied Biosystems®) according to the following conditions: initial denaturation at 95°C for 15 min, followed by 45 cycles of 30 s at 95°C, 60 s at 54°C and 45 s at 72°C. A final elongation step was held at 72°C for 15 min.

Successfully amplified PCR products were purified and sequenced for both strands at the High-Throughput Genomics Unit, Department of Genome Sciences, University of Washington (http://www.htseq.org/). Sequence trace files were checked and aligned by DNASTAR 5.0 package (DNASTAR Inc., Madison, WI) and aligned by software Mega version 5 (Tamura *et al.* 2011). All sequences generated in this work are available at GenBank (accession numbers KJ399477–KJ399523).

Available mtDNA control region sequences for this species were downloaded from GenBank and added to the obtained dataset (Table 1). Diversity measures, namely, haplotype and nucleotide diversities were calculated by DnaSP 5.10 software (Barcelona, Spain) (Librado & Rozas 2009) and software Network 4.6 (Bandelt 1999) was used to draw a median-joining network of haplotypes.

Table 1 - Haplotypes and sequenced base pairs (bp) obtained from DNA analysis of fecal samples Grevy´s zebra (*Equus grevyi*) collected from Ethiopia

Samples	Country	Location	Haplotype	Sequence length(bp)	Reference Paper
EG002, EG006, EG007, EG009, EG010,EG013, EG015, EG016, EG017, EG018, EG028, EG036, EG041, EG043, EG047	Ethiopia	Alledeghi	H2	329	Current
EG012, EG014, EG037, EG048	Ethiopia	Alledeghi	H3	329	Current
EG179, EG180, EG182, EG183, EG186, EG190, EG192, EG193, EG194, EG196, EG197, EG198, EG200, EG201, EG202, EG203, EG204, EG205, EG206, EG207, EG208 EG209, EG210, EG211	Ethiopia	Alledeghi	H1	329	Current
GQ176428, GQ176429, GQ176430, GQ176432	Kenya	Laikipia	H1	329	Cordingley *et al.* (2009)
AF220928, AF220930	Unknown	S. Diego Zoo	H1	329	Oakenfull *et al.* (2000)

Results and discussion

Twenty-five sequences were obtained for the Sarite population and twenty-one for the Alledeghi population. Four sequences were published in a study by Cordingley *et al.* (2009) and belonged to individuals sampled in the Laikipia region of Kenya. The remaining two sequences belonged to captive individuals and there was no information about their geographical origin. Three haplotypes were defined by two mutational steps and named H1, H2 and H3 (Fig. 2).

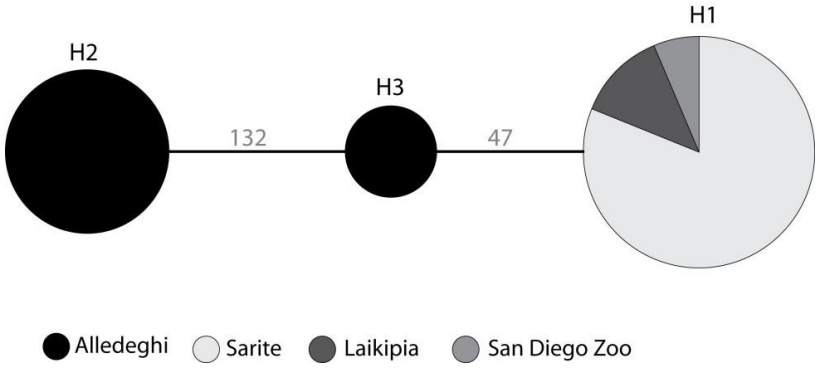


Figure 2 - Median-joining network of three Grévyi's zebra (*Equus grevyi*) haplotypes.

Haplotype H1 was the most frequent and it was shared by the population in southern Ethiopia (Sarite) and Kenya reflecting their geographical proximity. Two previously unpublished haplotypes were identified in the northern population of Alledeghi which were not shared with individuals that belonged to either Kenya or Sarite. Obtained results for haplotype and nucleotide diversities were low (Table 2). Values for the northern population of Alledeghi (h =0.381 and π = 0.00116) were higher than those found in the Sarite population (h = 0 and π = 0) where only one haplotype was found.

Table 2 - Genetic diversity parameters obtained from mtDNA control region sequencing for several zebra species

Species	Population	Location	N	h	π	Ref. paper
Equus grevyi	Ethiopia	Alledeghi	21	0.381	0.001	Current study
		Sarite	25	0	0	Current study
	Kenya	Laikipia	4	0	0	Cordingley et al., 2009
	San Diego Zoo	*	2	0	0	Oakenfull et al., 2000
Equus zebra	Namibia	Gamsberg	11	0.894	0.017	Moodley & Harley, 2005
		Kamanjab	13	0.962	0.019	"
		Cradock	10	0	0	"
		Kammanassie	11	0.545	0.006	"
Equus quagga	Kenya	Masai Mara	19	0.98	0.030	Lorenzen et al., 2008
	Tanzania	Maswa	20	0.98	0.032	"
		Burko	20	0.88	0.026	"
		Ikiri-Rugwa	23	0.92	0.016	
	Zambia	Kasama	10	0.96	0.027	"
		Lochri var South	11	1.00	0.016	"
	Namibia	Etosha	24	0.96	0.025	"

Zebra species (*Equus quagga*, *Equus zebra* and *Equus grevyi*) have mostly allopatric distribution in the African continent that reflects their physiology, feeding ecology and distinct ecological requirements (Bauer *et al.* 1994). Plain zebras (*Equus quagga*) are widely distributed from Ethiopia to southern Africa and are dominant members of savannah ecosystems (Lorenzen *et al.* 2008). Lorenzen *et al.* (2008) analyses of the mitochondrial DNA control region determined that the morphological variation of Plains zebra was consistent with higher levels of genetic variation (Table 2). Genetic diversity parameters were obtained from mtDNA control region sequencing for several zebra species [(π) nucleotide diversity and (h) haplotype diversity; Table 2].

In southern Africa, the Mountain zebra (*Equus zebra*) inhabits more rocky terrain and may have a comparatively lower dispersal potential. Hartmann's Mountain zebra (*E. z. hartmannae*) mainly occurs in Namibia and the Cape Mountain zebra persists in South Africa (*E. z. zebra*). Both subspecies have experienced population fluctuations / reductions. The Hartmann mountain zebra, although it may have experienced an ephemeral bottleneck, has recovered to an estimated 25,000 plus individuals (Moehlman *et al.* 2008) and this subspecies has a reasonable level of genetic diversity (Table 2) (Moodley & Harley 2005). By contrast, during the past 300 years, the Cape mountain zebra populations had a very severe and prolonged bottleneck and by the 1950's the population was less than 100 individuals (Novellie 2008). In the 1960's and 1970's conservation actions and reintroductions led to a recovery of the subspecies such that the population is now over 1500 individuals. The small number of founders appears to have had consequences, and the overall diversity level is low (Table 2). Some Cape Mountain zebra populations i.e. Cradock, Kammanassie, and Gamka, have only one haplotype and low genetic diversity.

Comparison of the genetic diversity values for these three species of zebra (Table 2) observed at the same genetic marker (mtDNA control region) indicate that, given the current information, some Grevy's zebra populations exhibit very low levels of nucleotide diversity. This raises concerns as to their genetic resilience for their long-term survival.

The Alledeghi Grevy's zebra population is at the northern extreme of the species' range and is geographically isolated. It retains what appears to be unique mitochondrial DNA and haplotypes.

The Grevy's zebra in Ethiopia is an important flagship species for the Alledeghi Plains grassland. In addition, this population appears to have a unique genetic makeup and strong support is needed for a community based conservation programme in this multiple use area. The current analyses are based only on mtDNA control region sequences and this information needs to be combined with microsatellite data as a next step in understanding the population genetics of the Alledeghi Grevy's zebra population.

Additional population genetic studies are ongoing to determine whether the Alledeghi and Sarite populations are still connected (via male dispersal) and, if the southern Sarite population currently has genetic exchange with the northern Kenya populations. Further work is needed on Grevy's zebra population genetics throughout its range in order to understand its genetic diversity distribution and dynamics.

References

Bandelt H.-J., Forster, P., Rohl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37-48.

Bauer I.E., McMorrow J. & Yalden D.W. (1994) The historic ranges of three equid species in north-east Africa: a quantitative comparison of environmental tolerances *Journal of Biogeography* **21**, 169-82.

Beja-Pereira A., Oliveira R., Alves P.C., Schwartz M.K. & Luikart G. (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. *Mol Ecol Resour* **9**, 1279-301.

Broquet T., Ménard N. & Petit E. (2007) Noninvasive population genetics: a review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. *Conservation Genetics* **8**, 249-60.

Cordingley J.E., Sundaresan S.R., Fischhoff I.R., Shapiro B., Ruskey J. & Rubenstein D.I. (2009) Is the endangered Grevy's zebra threatened by hybridization? *Animal Conservation* **12**, 505-13.

Frankham R., Ballou, J.D. and Briscoe, D.A. (2002) *Introduction to Conservation Genetics*. University Press, Cambridge, UK.

Kebede F., Bekele A., Moehlman P.D. & Evangelista P.H. (2012) Endangered Grevy's zebra in the Alledeghi Wildlife Reserve, Ethiopia: species distribution modeling for the determination of optimum habitat. *Endangered Species Research* **17**, 237-44.

Librado P. & Rozas J. (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-2.

Lorenzen E.D., Arctander P. & Siegismund H.R. (2008) High variation and very low differentiation in wide ranging plains zebra (*Equus quagga*): insights from mtDNA and microsatellites. *Molecular Ecology* **17**, 2812-24.

Moehlman P.D., Rubenstein P. & Kebede F. (2008) *Equus grevyi*. URL http:\\www.iucnredlist.org.

Moodley Y. & Harley E.H. (2005) Population structuring in mountain zebras (*Equus zebra*): The molecular consequences of divergent demographic histories. *Conservation Genetics* **6**, 953-68.

Moritz C. (1994) Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecology & Evolution* **9**, 373-5.

Novellie P. (2008) *Equus zebra*. URL http:\\www.iucnredlist.org.

Oakenfull E.A., Lim H. & Ryder O. (2000) A survey of equid mitochondrial DNA: Implications for the evolution, genetic diversity and conservation of *Equus*. *Conservation Genetics* **1**, 341-55.

Taberlet P., Griffin S., Goossens B., Questiau S., Manceau V., Escaravage N., Waits L.P. & Bouvet J. (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Res* **24**, 3189-94.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M. & Kumar S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*.

Williams S.D. (2002) Status and Action Plan for Grevy´s Zebra (Equus grevyi). In: *Equids: Zebras, Asses and Horses. Status Survey and Conservation Action Plan* (ed. by Moehlman PD), pp. 11-27. IUCN, Gland, Switzerland.

Chapter VI

General Discussion

6.1 General Discussion

This thesis represents the first assessment on the evolutionary history of the African wild (*Equus africanus*). Despite its rich evolutionary history, the African wild ass is currently facing an extremely high risk of extinction with few individuals subsisting in extreme habitats in Ethiopia and Eritrea. In fact, from the three proposed African wild ass subspecies, only the Somali wild ass is confirmed to still subsist. The Atlas wild ass became extinct in historic times (Van Bemmel 1972) and the Nubian wild ass is currently extremely rare or even extinct.

To loose this species would mean losing the possibility of studying the domestication process from a unique perspective, by comparing ancestral (African wild ass) and descendent (domestic donkeys) genomes. It would also mean, losing an important part of the *Equus* genus inheritance, as it has happened along the course of times, with many other *Equus* species that became extinct, such as the direct ancestral of the domestic horse.

Up to now, there was no published study that provided information about the levels of genetic diversity in natural extant populations of ass-like equids. On this respect, we have calculated and compared levels of genetic diversity, in African and Asiatic wild ass populations (Chapter II, **Article 1**). From all, the Kalamaili population of *Equus hemionus hemionus* in China, showed the highest nucleotide diversity values from all populations and ass-like equid species. Both *Equus hemionus onager* populations in Iran presented low diversity values which is concerning given that these are the only remaining populations of this subspecies that has been undergoing a severe population reduction. Finally, within the African wild ass species, the Eritrean population was the one presenting the highest genetic diversity values.

Distinct phylogeographic patterns were found between Asiatic and African wild ass species. The African wild ass populations in Ethiopia and Eritrea shared three out of the four obtained mtDNA haplotypes showing no geographic structure, however there was a clear differentiation between Somali wild ass haplotypes and those published from Nubian wild ass haplotypes (Beja-Pereira *et al.* 2004a; Kimura *et al.* 2011). The *Equus hemionus* haplotype network showed no haplotype sharing between populations in Iran and China, however, there was no marked differentiation among *E.h.hemionus and E.h.onager, with one E.hemionus onager* haplotype grouping closer to *E.hemionus hemionus haplotypes,* what might be a consequence of a recent and ongoing process of differentiation between *E. h hemionus* and *E. h. onager* subspecies. *Equus kiang* haplotype network revealed a marked east/west geographical structuring (Fig. 5C), with the most extreme populations being represented by the two most divergent haplotypes. This different phylogeographic patterns among ass-like equids may reflect their evolutionary history and that of

the different geographic features of the habitats and the level and temporal length of fragmentation/isolation among extant populations.

Current taxonomic status of species/subspecies was not supported by the analyses of molecular variance. Grouping of geographically close populations of *Equus kiang* and *Equus hemionus hemionus* in China, produced higher "among group" variation, then grouping taxonomically close subspecies of *Equus hemionus hemionus* (China) and *Equus hemionus onager* (Iran). This results were further supported by the phylogenetic analyses of cytochrome b (Cyt *b*) haplotypes, which revealed a clear pattern of differentiation between *E. h. onager* and the cluster incorporating *E. kiang* and *E. h. hemionus* haplotypes.

We have calculated time to the most recent common ancestor (TMRCA) using newly obtained cytochrome b sequences of Asiatic and African asses. The obtained results that this group shared a common ancestor approximately 2.3 Mya (95% CI; 1.4-3.2 Mya), what is remarkably in line with the timeframe is obtained from the fossil record for the putative ancestor of the wild ass branch.

The *Equus* genus had its origin the North America, approximately 4 million years ago (Orlando *et al.* 2013), and dispersed to Euroasia before the end of the Pliocene. Given this scenario it is probable that wild asses co-existed with early horses in North America, before dispersion to the Old World occurred. Despite being subject to "similar" climatic fluctuations during the Pleistocene, the impact of this changes in the demography of Asiatic and African wild asses had very different outcomes. Asiatic wild asses suffered a detectable population decline approximately 25,000 years ago, coinciding with the time of the last maximum glacial (LGM), between 19,000 and 26,000 years ago, while African wild asses maintained lower but more stable effective population sizes, over time.

One of the main advantages of the genome sequencing projects of the economic important species, such as the horse, is the number of polymorphic markers that are published and available. As horses and asses are relative close species, belonging to the same genus, it is possible to conduct cross-species studies to identify microsatellite loci that can be used in population genetics studies of the wild and domestic asses. To identify horse polymorphic microsatellites suitable to be used in wild asses population genetics assessments (Chapter III; **Article 2**) we have tested and characterize twenty-five previously published horse microsatellite and amplified twenty-two captive African wild ass samples of mixed origin (e.g., blood, tissue and feces). Null allele rates of the chosen microsatellites, were overall low, with the exception of one locus, meaning that alleles were reliably being amplified and detected, and therefore that obtained genotypes were reliable. Levels of heterozygosity and allelic richness were moderately high and

so this microsatellite set was considered adequate to be used at a wider level, in population genetics studies of the African wild ass. As expected, our results showed that horse isolated microsatellite markers were adaptable for other Equid species, namely the African wild ass.

Non-invasive samples (e.g., feces) were used as a source of DNA from wild individuals from the three remaining populations of African wild ass (Chapter III, **Article 3**), to genotype a subset of 10 of those microsatellites described in Article 2 (Rosenbom *et al.* 2011). Our data pointed to a weak geographic structure of African wild ass populations, although the Afdera population, in Ethiopia, showed a slightly higher degree of differentiation.

Both Ethiopian and Eritrean populations presented very small effective population sizes, what can be an added problem, given the potential for the enhanced effects of inbreeding and genetic drift. Nonetheless, the overall estimates of effective population sizes for studied populations seem to be in good agreement with the overall proposed census number of approximately 600 individuals in Eritrea and Ethiopia (Moehlman 2008b), given the reference value for the *Ne/N* ratio of approximately 0.11 (Frankham 2003b; Frankham 2009; Frankham 2002).

Previously reported domestic mtDNA haplotypes were detected in morphologically identified African wild asses, both in Ethiopia and in Eritrea, however only one of this individuals presented a pattern of admixture consistent with simulated F1 hybrids. Such fact is consistent with the existence of sporadic and geographically limited events of hybridization between the wild and the domestic forms, and thus, minimizes any concerns about the integrity of the African wild ass genetic pool.

Another important aspect that can help in the conservation of this species is to test whether its small *N*e populations are connected or not. Indeed, our migration analyses points towards the existence of both historical and contemporary gene-flow among studied African wild ass populations. This can be seen as a positive aspect as gene flow will certainly help mitigate the enhanced risk of drift and inbreeding, as long as human-mediated changes to their habitat do not isolate these populations. Thus, operational migratory cross boundaries corridors must be maintained in order to prevent the increase of population structuring and prevent further isolation of extant populations.

The African wild ass has long been hypothesized as the most probable ancestor of the domestic donkey. The existence of three African wild ass subspecies has been proposed, according to the historical geographic distribution of the species. The Nubian wild asses (*Equus africanus africanus*), distributed in Sudan and Northern Eritrea, the Somali wild ass (*Equus africanus*

somaliensis), from southern Eritrea, Ethiopia and Somalia and the Atlas wild ass (*Equus africanus 'atlanticus'*), once confined to the northwestern part of the continent and extinct in early historic times. Which of the proposed subspecies were in fact involved in the process of donkey domestication has been in the center of debate in studies concerning donkey domestication (Beja-Pereira *et al.* 2004a). A vast mtDNA survey on Old World donkeys, encompassing 52 countries, demonstrated the existence of two distinct mitochondrial haplogroups, termed Clades 1 and 2. The fact that Nubian wild ass haplotypes clustered within the Nubian clade (Clade 1), led to the acceptance of this subspecies as the putative ancestor of Clade 1 donkeys. Clade 2 ancestry however has been more challenging to unravel. The most probable candidate as the ancestor of Clade 2 donkeys, the Somali wild ass, was not confirmed, hence mtDNA control region haplotypes of the Somali wild ass failed to cluster clearly within the wider variation of Clade 2 donkeys. It was than hypothized that a relative of the Somali wild ass, probably already extinct, would be the most probable ancestor of Clade 2 donkeys. In order to further support the currently accepted hypotheses of Nubian wild asses being the ancestors of Clade 1 domestic donkeys and a relative of the Somali wild ass being the ancestor of donkeys of Clade 2, a comprehensive study including ancient archaeological and historic museum samples, as well as contemporary Somali wild ass samples was designed and resulted in **Article 4**, in Chapter III.

Results on this work showed that only three out of twelve ancient samples, with ages that varied from 7000 BC to 117 BC, were successfully sequenced for the mitochondrial control region. One this samples had the maternal genetic signature of a horse and was not further analyzed. The two remaining samples, were collected in the Uan Muhuggiag rock shelter in Lybia, and produced a sequence that fell in Clade 1, supporting an ancestral role for the Nubian wild ass within Clade 1. Two samples, dated at approximately 4900 BC, collected in Yemen, failed to produce results. This fact was particularly disappointing given the impossibility of verifying if this individuals, which were clearly pre-domestication, represented in fact a different African wild ass lineage, corresponding to a possible extinct population.

Nine historical samples (1886 AD - 1939 AD), belonged to animals phenotypically and geographically identified as Nubian wild ass. Eight of these samples produced five different sequences that fell inside the variation of Clade 1, being that four of these corresponded to a haplotype also find in domestic donkeys. A sequence from the Atbara region in Sudan, exactly matched a modern sequence from eastern Sudan, tentatively identified as Nubian wild ass, suggesting that Nubian wild ass maternal lineages survived at least until the last decade in eastern

Sudan. A ninth historic sample attributed to Nubian wild ass on phenotypic grounds had a sequence identical to a haplotype found in domestic donkeys of Clade 2.

The only historical Somali wild ass sample belonged to a specimen from Berbera, Somalia and was collected around 1886. It produced a sequence identical to one of the obtained modern Somali wild ass haplotypes, from Eritrea and Ethiopia. This results demonstrates a historical continuity in the mitochondrial variability of the Somali wild ass, within the region over the last 120 years.

Modern Somali wild ass specimens produced four new haplotypes that fell in the same clade as previous Somali wild ass specimens. This clade is well separated from domestic donkey Clades 1 and 2 and clearly not ancestral to either clade. Haplotype diversity in the African wild ass clade was lower than in both domestic donkey Clades, suggesting that the genetic variability in contemporary Somali wild ass populations is low. Obtained African wild ass haplotypes were found in Eritrea and Ethiopia, presenting no geographic structuring.

Overall results of this work enabled us to confirm the role of the Nubian wild ass as an ancestor of Clade 1 domestic donkeys. In fact, Nubian wild ass and Clade 1 domestic donkeys are basically indistinguishable on the basis of mtDNA, sharing some of the obtained haplotypes. These shared haplotypes most probably represent survival of the originally domesticated maternal haplotypes in the wild populations. It was also possible to reevaluate the historic distribution of the Nubian wild ass, which is now believed to range from northern Sudan and Eritrea, as far as west as the central Sahara.

The role of the s Somali wild ass in donkey domestication was also further clarified in this work. Despite the large sample size, only four new haplotypes were obtained among Somali wild ass sample, making it unlikely that unsampled lineages exist and, thus, making the Somali wild ass a less probable candidate to Clade 2 ancestry. Alternative hypothesis for Clade 2 ancestry are an unsampled Nubian-like wild ass, and that would mean that both Clades have identical ancestry, or an already extinct population.

Besides this two African hypotheses for the domestication of the donkey, the identification of putative *Equus africanus* remains at sites in the Levant and the Arabian Peninsula, raised the possibility of a west Asian domestication of the donkey (Meadow R.H. 1991; Uerpmann 1987; Zeder 1986). Scarce genetic data, has provided support to Northeastern Africa as the most probable geographic area for donkey domestication. Genetic diversity assessment of control region sequences from domestic donkeys across the Old World, has shown that both domestic

donkey Clades, present significant higher genetic diversity in that region of the world (Beja-Pereira *et al.* 2004a).

The genetic analysis of eight donkey populations from within the three hypothesized centers of origin of the domestic donkey - Northeastern Africa (Ethiopia, Sudan, Egypt), Arabian Peninsula (Oman, Yemen) and Near East (Syria, Turkey, Jordan), showed that the highest observed Genetic diversity values (namely allelic richness and private allelic richness) were found at the Northeastern Africa and the Arabian Peninsula (Chapter IV, **Article 5**). At a finer geographic scale the Sudanese population was the one displaying the highest levels for calculated genetic diversity parameters (H_O, H_E, uH_E and Ar) (Table 2). Nonetheless, Yemen possessed the highest values for private allelic richness (PAr = 0.55) and the second highest values for allelic richness (Ar = 5.78). Overall, Sudan and Yemen stand out, among countries in putative centers of origin, as possessing the highest values for analyzed diversity parameters, indicating these countries in particular as potential sources of unique genetic diversity.

Allelic patterns results were also enlightening. The vast majority of found putative private alleles were distributed in Northeastern Africa and the Arabian Peninsula, however, when looking at alleles at frequencies between 2% and 5%, we were able to identify alleles found in the Arabian Peninsula, as occupying marginal positions in allelic frequency plots.

Obtained results have pinpointed two geographical areas with high diversity values, corresponding to two potential centers of origin of the domestic donkey. In particular, the high genetic diversity found in Sudan is consistent with the historical distribution of the putative ancestor of Clade 2 donkeys – the Nubian wild ass – which was historically distributed in Sudan and Northern Eritrea. The fact that Yemen presented the highest levels of unique diversity, measured as private allelic richness, among analyzed countries, raises the possibility of a domestication event occurring in this geographic region. The presence of African wild ass or early domesticated donkeys in a site in Yemen is well-documented and considered reliable among archaeologists (Marshall 2007). Excavated bones were dated at approximately 7770 ± 95 BP, clearly predating the proposed dates for donkey domestication of 5000 years ago (Beja-Pereira *et al.* 2004a; Kimura *et al.* 2011). This fact makes plausible the existence of a wild population, in the Arabian Peninsula, during ancient times. This scenario favors a possible contribution of a wild population from the Arabian Peninsula to the gene pool of the domestic donkey, as suggested by archaeological data (Cattani & Bokonyi 2002).

Grevyi´s zebra shares similar ecological and habitat features with the African wild ass and it also presents a high level of endangerment. These make Grevyi's zebra an interesting species to

compare with the African wild ass in terms of genetic diversity and population structure (Chapter V, **Article 6**). The mtDNA sequencing analysis showed striking low levels of diversity for the northern population of Alledeghi (h = 0.381 and π =0.00116), even low ones found in the Sarite population (h = 0 and π = 0). Just three haplotypes, defined by two mutational steps, were observed. The comparison of such low values with those obtained in other species of zebra, namely the Plains zebra (*Equus quagga*) and the Mountain zebra (*Equus zebra*), using the same genetic marker (mtDNA control region) indicated that Ethiopian Grevy's zebra populations exhibit very low levels of nucleotide diversity, which raises concerns as to their genetic resilience for long-term survival of these populations.

6.2 Future perspectives

My personal goal is to pursue my research work in the areas of Equid evolutionary history, conservation and population genetics, with a special focus on wild ass species. In order to do so, it is absolutely mandatory to develop a modern, genome-wide molecular approach, using non-invasive samples from a reasonable number of individuals per population.

The increase of publically available genomic data for many species and the decrease on costs associated with modern sequencing techniques, will allow in the future for the wide use of genomic data in many study areas. In the particular case of equids, genomic data has already been used for recalibrating *Equus* evolution (Orlando *et al.* 2013), with valuable insights being achieved. However, much more could be achieved in the next years, by exploring variation in natural ass-like equid populations at a genomic scale.

Genomic based studies on the evolutionary and demographic history of wild ass species will allow for more robust inferences and overall improvement on the knowledge of the evolutionary history of this unique group of animals.

The first essential step will be the development of a new methodological approach that will allow the retrieval of genomic data from non-invasive samples, namely fecal samples. The challenge will be to achieve high quality genomic data from low quality/quantity DNA, as is the case of DNA extracted from feces or hair.

Obtained genomic data will also be used in order to investigate adaptive divergence in wild ass populations, hence understanding the genetic basis of adaptation to different environments represents an important goal in evolutionary biology.

It is also my objective to continue my research in donkey domestication. Newly obtained evidence about the possible contribution of the Arabian Peninsula in the donkey domestication process, has made this geographic region a particular important region to be sampled in future domestication studies. Additionally, it would be important to obtain historical and ancient samples of putative African wild asses or early domestic donkeys from the Arabian Peninsula in order to detect if any new African wild ass lineages would be recovered in this region. The use of genomic data in ancient Equid samples has already been achieved (Orlando *et al.* 2013) and we can now be more positive about the successful use of these modern approaches in order to improve our knowledge in study areas such as animal domestication.

6.3 References

Beja-Pereira A., England P.R., Ferrand N., Jordan S., Bakhiet A.O., Abdalla M.A., Mashkour M., Jordana J., Taberlet P. & Luikart G. (2004) African origins of the domestic donkey. *Science* **304**, 1781.

Cattani M, Bokonyi S (2002) Ash-Shumah: an early Holocene settlement of desert hunters and mangrove foragers in the Yemeni Tihamah In: Essay on the late Prehistory of the Arabian Peninsula (eds. Cleuziou S, Tosi M, Zarins J), pp. 31-53. Instituto Italiano per l'Afirca e l'Oriente, Rome.

Frankham R (2003) Genetics and conservation biology. *Comptes Rendus Biologie* **326 Suppl 1**, S22-29.

Frankham R (2009) Effective population size/adult population size ratios in wildlife: a review. *Genetical Research* **66**, 95.

Frankham R, Ballou, J.D. and Briscoe, DA (2002) Introduction to Conservation Genetics University Press, Cambridge, UK.

Kimura B, Marshall FB, Chen S, et al. (2011) Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication. *Proceedings of the Royal Society B* **278**, 50-57.

Marshall F (2007) Rethinking Agriculture: Archeological and Ethnoarcheological perspectives. Left Coast Press, Inc., Walnut Creek, California, USA.

Meadow RH, Uerpman H-P (1991) Equids in the Ancient World Ludwig Reichert Verlag, Wiesbaden.

Moehlman PD, Yohannes, H., Teclai, R. & Kebede, F. (2008) *Equus africanus*. In: IUCN Red List of Threatened Species. Version 2013.2.A. URL http://www.iucnredlist.org.

Orlando L., Ginolhac A., Zhang G., Froese D., Albrechtsen A., Stiller M., Schubert M., Cappellini E., Petersen B., Moltke I., Johnson P.L.F., Fumagalli M., Vilstrup J.T., Raghavan M., Korneliussen T., Malaspinas A.-S., Vogt J., Szklarczyk D., Kelstrup C.D., Vinther J., Dolocan A., Stenderup J., Velazquez A.M.V., Cahill J., Rasmussen M., Wang X., Min J., Zazula G.D., Seguin-Orlando A., Mortensen C., Magnussen K., Thompson J.F., Weinstock J., Gregersen K., Roed K.H., Eisenmann V., Rubin C.J., Miller D.C., Antczak

D.F., Bertelsen M.F., Brunak S., Al-Rasheid K.A.S., Ryder O., Andersson L., Mundy J., Krogh A., Gilbert M.T.P., Kjaer K., Sicheritz-Ponten T., Jensen L.J., Olsen J.V., Hofreiter M., Nielsen R., Shapiro B., Wang J. & Willerslev E. (2013) Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**.

Rosenbom S, Costa V, Steck B, Moehlman P, Beja-Pereira A (2011) Cross-species genetic markers: a useful tool to study the world's most threatened wild equid—*Equus africanus. European Journal of Wildlife Research* **58**, 609-613.

Uerpmann H-P (1987) The Ancient Distribution of Ungulates Mammals in the Middle East: In Equids in the Ancient World (eds R.H Meadow & H-P Uerpmann) Ludwig Reichert Verlag, Wiesbaden.

Van Bemmel A.C.V. (1972) Some remarks on the African wild ass. *Zoologische Mededelingen* **47**.

Zeder MA (1986) The equid remains from Tal-e-Malyan, southern Iran. In: Equids in the Ancient World (ed. RH Meadow & Uerpann H-P) Ludwig Reichert Verlag, Wiesbaden.